



**António Manuel Duarte Cunha Azinheira da
Cruz**

Licenciado em Ciências de Engenharia e Gestão Industrial

**Utilização de cartas de controlo para
monitorização em tempo real da
qualidade da água numa rede de
distribuição**

Dissertação para obtenção do Grau de Mestre em
Engenharia e Gestão Industrial

Orientador: Prof^a. Doutora Ana Sofia Leonardo Vilela de
Matos, Professora Auxiliar, FCT-UNL

Co-orientador: Prof^o. Doutor Daniel Cardoso Vaz,
Professor Auxiliar, FCT-UNL

Júri:

Presidente: Prof. Doutora Virgínia Helena Arimateia de Campos Machado

Arguente: Prof. Doutora Teresa Paula Costa Azinheira Oliveira

Vogais: Prof^a. Doutora Ana Sofia Leonardo Vilela de Matos

Prof^o. Doutor Daniel Cardoso Vaz

Mestre Paulo Nico Casimiro

Utilização de cartas de controlo para monitorização em tempo real da qualidade da água numa rede de distribuição

Copyright © António Cruz, FCT/UNL e UNL

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Agradecimentos

“You can, you should, and if you’re brave enough to start, you will.”

Esta dissertação nasce como resultado do apoio e dedicação de muitas pessoas, às quais passo a exprimir o meu agradecimento mais sincero.

Em primeiro lugar, à minha orientadora, a Professora Ana Sofia Matos, pela constante simpatia que demonstrou desde o primeiro ano do meu curso, pela paciência eterna que demonstrou, por todo o apoio e disponibilidade, sem as quais a presente dissertação nunca existiria.

Ao meu co-orientador, o Professor Daniel Vaz, que foi capaz de iluminar o caminho nos momentos mais difíceis deste estudo com a sua paciência, conhecimentos e experiência, sem os quais este trabalho nunca poderia ter qualidade.

Aos Serviços Municipalizados de Água e Saneamento de Almada (SMAS de Almada), especialmente na pessoa do Engenheiro Paulo Nico, cuja disponibilidade, interesse e entrega ao projecto foram indispensáveis para a realização do mesmo.

Ainda do SMAS de Almada, ao Engenheiro Paulo Gonçalves, cuja disponibilidade, paciência e interesse no projecto contribuíram largamente à execução do mesmo.

Um especial agradecimento a todos os colegas e amigos que fiz na minha passagem pela FCT-UNL. Todos, sem excepção, me apoiaram e acompanharam ao longo do curso, me fizeram crescer enquanto estudante e pessoa, e os momentos que vivemos nunca serão esquecidos.

À Sofia Pedrosa, pessoa de valor inestimável, cujo apoio foi indispensável todos estes anos na viagem repleta de altos e baixos que foi esta passagem pela universidade.

Aos meus avós e tia, por tudo o que fizeram ao longo da minha vida para que hoje pudesse cá estar.

Por último, mas o mais importante de todos, ao meu pai, sem o qual não seria a pessoa que sou, e sem o qual nada disto teria acontecido.

A todos os que nomeei, o meu mais sincero obrigado!

Sumário

A água é um recurso crucial ao ser humano, e como tal a sua gestão e distribuição deve ser um exemplo da excelência e eficácia em todos os seus aspectos. No Mundo 11% da população ainda não tem acesso a água potável. Mesmo nas sociedades mais desenvolvidas, onde o acesso a água potável está melhor assegurado, esta nem sempre tem qualidade excelente. Importa pois continuar a realizar esforços no sentido de garantir que todos os cidadãos usufruam de água potável nas suas habitações com a qualidade que lhes é devida. Nesse seguimento surge a necessidade de complementar a gestão e distribuição da água potável com sistemas de informação em tempo real e técnicas estatísticas e de qualidade para utilizar as potencialidades do sistema de gestão de forma a garantir um controlo das características da qualidade ao longo da rede de distribuição.

Utilizando a condutividade da água como um dos indicadores da sua qualidade, nesta dissertação propõe-se desenvolver um modelo capaz de prever o valor dessa característica em função dos valores das características da água que entra numa rede de distribuição, dos caudais dessas mesmas entradas, do volume de água presente no reservatório de saída da rede e do tempo que a água demora a chegar a esse mesmo reservatório vinda de cada uma das entradas do sistema.

A partir dos dados reais da condutividade da água de cada uma das entradas, dos valores dos caudais e volumes de água obtidos a partir da telegestão do SMAS de Almada foi possível construir um modelo capaz de prever o valor da condutividade da água que chega ao reservatório de saída, numa rede composta por quatro inputs e um reservatório de saída (output).

Construído o modelo de previsão, procedeu-se ao controlo dos inputs e outputs do sistema, através de cartas de controlo. As quatro entradas da rede analisada dividem-se em três captações subterrâneas e um reservatório idêntico ao reservatório de saída. Às captações subterrâneas aplicaram-se directamente cartas de controlo Q para controlar o valor da condutividade que seria utilizado no modelo de previsão, sendo que o reservatório de entrada sofreu o mesmo tratamento mas apenas aos valores mais recentes, devido à sua grande variabilidade.

Controlados os inputs de água e a previsão, torna-se necessário comparar a previsão da condutividade ao valor real, através da modelação da série temporal obtida com recurso a modelos estatísticos e controlo estatístico do processo. Aplicando a fase I com cartas de controlo de observações individuais e amplitudes móveis e a fase II com cartas EWMA (*Exponentially Weighed Moving Average*) de resíduos e MCEWMA (*Moving centre-line exponentially weighed moving average*) foi possível observar que a previsão obtida do modelo tem qualidade e as cartas de controlo da mesma detectam também eventuais desvios da previsão face ao valor real.

O modelo foi desenvolvido tendo como base um pequeno troço da rede total de distribuição de água. O modelo, validado através do *software Matlab*, pode ser estendido a toda a rede e irá permitir, com as cartas de entrada, detectar contaminações nas captações ou outro tipo de anomalias, e com as cartas à saída detectar valores elevados de condutividade que possam ser devidos a contaminações entre a captação e o reservatório, ou ao somatório de condutividades elevadas provenientes das diferentes captações. A metodologia apresentada também se poderá aplicar a outro tipo de fluidos, em que a qualidade possa ser aferida pela condutividade ou outra propriedade passível de ser medida on-line.

O SMAS de Almada beneficia deste estudo na medida em que ganha capacidade de reacção face a eventuais fontes de poluição da água que distribuem à população do Concelho.

Palavras-chave: Qualidade da água; Cartas de controlo; SPC para processos autocorrelacionados; Controlo de qualidade em redes de distribuição; Condutividade eléctrica.

Abstract

Water is a crucial resource for human beings, therefore its management and distribution has to be an example of excellence in all aspects. It is crucial to guarantee that all citizens have access to quality drinking water. Therefore the need arises to complement its management with real time information systems, statistical and quality techniques to use all the system's potential to assure quality control throughout the distribution network.

Using water conductivity as an indicator of water quality, in this dissertation a model is created that can predict the value of this indicator as a function of the characteristics of the water that enter any given distribution network, the flow of water coming from those entrances and the volume present in any given reservoir that serves as the exit of the network, as well as the time the water takes to reach the said exit.

Based on the real data of water conductivity in each one of the entrances (inputs), the water flow and volumes obtained from the SMAS' "telegestão" system it is possible to build a model that can predict the conductivity value at the exit of a network with four inputs and one exit reservoir (output).

Using the prediction model, the next step was the input control, using control charts. The four water entrances split between three underground water wells and one reservoir similar as the reservoir at the exit of the network. To the three underground water wells, Q charts were directly applied to control the conductivity value, while due to the extreme variability the reservoir input saw the same charts being applied, only to its more recent values.

After the input control, it becomes necessary to compare the forecast of the water conductivity at the exit of the network to its real value, through time series modeling obtained using statistical methods and statistical process controlling. Applying phase I with individual observation and moving range charts and phase II with residual EWMA (*Exponentially Weighed Moving Average*) and MCEWMA (*Moving centre-line exponentially weighed moving average*) charts it became possible to claim that the forecasted values generated by the model are solid when compared to the real values, and also that the charts stated above were capable to detect possible deviations of the forecasted values from the real values.

The prediction model was developed based on a small section of the SMAS water distribution network and validated using *Matlab software*. It can be extended throughout the entire network, and will allow, along with the control charts at the inputs to detect contaminations and other anomalies with water entering any network. With the control charts at the exit of the network it can detect high conductivity values, that can be a result of leaks or contaminations between the underground water wells and the reservoirs, or to the sum of high conductivity values coming from different underground wells. This methodology can also be applied to other fluids, where quality can also be measured from conductivity or other variable that can be measured on-line.

The SMAS of Almada will benefit from this study as they improve their capacity to react to possible pollution sources in the water they distribute through their networks.

Keywords: Water quality; Control charts; Autocorrelated process SPC; Quality control in distribution networks; Water conductivity.

Abreviaturas

AR	<i>Auto-regressive</i>
ARL	<i>Average run length</i>
CUSUM	<i>Cumulative sum</i>
DMA	Desvio médio absoluto
EQM	Erro quadratico médio
EWMA	<i>Exponentially weighed moving average</i>
FIR	<i>Fast initial response</i>
I	<i>Integrative</i>
JIT	<i>Just-in-time</i>
LC	Linha central
LIC	Linha inferior de controlo
LSC	Linha superior de controlo
MA	<i>Moving average</i>
MCEWMA	<i>Moving centre-line exponentially weighed moving average</i>
MR	<i>Moving range</i>
SMAS	Serviços municipalizados de água e saneamento
SPC	<i>Statistical process control</i>
KS	Kolmogorov-Smirnov

Simbologia

A	Área
A_t	Valor do erro amortecido no instante t
B^m	Operador de desfasamento de m unidades de tempo
c	Constante
C_p	Índice de capacidade potencial
C_{pk}	Índice de capacidade
C_{pki}	Índice de capacidade inferior
C_{pks}	Índice de capacidade superior
$COV(x,y)$	Covariância entre as variáveis x e y
D_3, D_4, d_2, c_4	Termos constantes das cartas de controlo
DAM_t	Valor do desvio médio absoluto no instante t
$E(Y)$	Valor esperado da variável Y
e_t	Erro ou desvio no instante t
E_t	Valor da estatística E no instante t
$F_{a,b}$	Função distribuição de Fischer, com a e b graus de liberdade no numerador e denominador, respectivamente
G_r	Distribuição t-student inversa para r graus de liberdade
j	Conduta ou segmento de conduta
k	Constante respeitante ao tipo de especificação
K	Constante da carta EWMA
L	Comprimento
n	Tamanho da amostra
Q_i	Índice capacidade inferior da carta Q

Q_r	Valor da estatística Q no instante r
Q_s	Índice capacidade superior da carta Q
Q_v	Caudal volúmico
r_{xy}	Coeficiente de correlação entre as variáveis x e y
S_x	Estimador do desvio padrão da variável x
T	Tempo
$(Ts)_t$	Valor da estatística tracking signal no instante t
$\text{Var}(Y), \sigma_Y^2$	Variância da variável Y
W_r	Valor da estatística W no instante r
\bar{Y}	Valor médio da variável Y
Y_t	Valor observado da variável Y no instante t
Z_α	Distribuição normal com um nível de significância α
α	Nível de significância
Φ^{-1}	Inverso da função distribuição normal
λ	Constante de amortecimento da carta EWMA
λ_1	Constante de amortecimento do erro amortecido
μ_Y	Valor esperado da variável Y
ϕ_i	i-ésimo coeficiente de regressão do modelo AR
$\phi(B)$	Função do modelo AR
ρ_k	Coeficiente de autocorrelação para uma lag k
δ	Constante multiplicadora da alteração da média a detectar
θ_i	i-ésimo coeficiente de regressão do modelo MA
∇Y_t	Operador diferencial

Δ	Alteração da média a detectar
$\theta(B)$	Função do modelo MA
v	Estimador da variância do coeficiente de autocorrelação

Índice geral

Agradecimentos.....	v
Sumário.....	vii
Abstract.....	ix
Abreviaturas.....	xi
Simbologia.....	xiii
Índice geral.....	xvii
Índice de figuras.....	xxi
Índice de tabelas.....	xxiii

Capítulo 1 – Introdução	1
1.1 Enquadramento e motivação	1
1.2 Objectivos	2
1.3 Estrutura da dissertação.....	2
Capítulo 2- SMAS de Almada	5
2.1 A entidade gestora.....	5
2.1.1 Estratégia da empresa/entidade	5
2.1.2 Gestão e organograma da empresa/entidade	5
2.2 Rede de abastecimento de água.....	6
2.3 Telegestão	8
2.4 Controlo da qualidade da água nos SMAS de Almada	8
Capítulo 3 – Fundamentos Teóricos	11
3.1 A condutividade da água	11
3.2 A qualidade	12
3.3 As sete ferramentas básicas da qualidade.....	14
3.4 Análise de séries temporais	14
3.4.1 Correlação de dados	15
3.4.2 Autocorrelação de dados	15
3.4.3 Autocorrelação parcial	18
3.4.4 Modelos ARIMA de Box-Jenkins.....	18
3.4.5 Estimação dos parâmetros p, d, q do modelo ARIMA.....	20
3.4.6 Média e variância de processos ARIMA.....	21
3.4.7 Teste à normalidade de uma série de dados	22
3.5 Controlo estatístico do processo.....	23
3.5.1 Variabilidade e causas de variação.....	25
3.5.2 Detecção de causas especiais de variação	25

3.5.3	Capacidade dos processos	26
3.5.4	Fase I e Fase II do SPC	27
3.5.5	Medida de desempenho das cartas de controlo	28
3.5.6	Carta X e MR	28
3.6	Controlo de pequenas produções.....	30
3.6.1	<i>Short Runs</i>	30
3.6.2	Cartas Q.....	30
3.6.3	Limites de controlo para as cartas Q	32
3.6.4	Análise da capacidade do processo com as cartas Q.....	32
3.7	Cartas CUSUM e EWMA	33
3.7.1	Cartas de controlo especiais	33
3.7.2	Carta EWMA de resíduos	35
3.7.3	Carta MCEWMA	35
3.8	Utilização do SPC no controlo da qualidade da água.....	37
3.9	Algumas noções de dinâmica de fluidos	38
3.10	Qualidade da previsão	39
3.11	Análise dos dados e estudo de <i>outliers</i>	40
Capítulo 4 – Metodologia.....		43
4.1	Introdução	43
4.2	Análise dos dados de entrada	43
4.3	Construção do modelo de previsão temporal	45
4.4	Modelo não temporal	49
4.5	Modelo híbrido.....	49
4.6	Controlo e monitorização da qualidade.....	50
Capítulo 5 – Discussão de Resultados.....		51
5.1	Modelo temporal	51
5.2	Modelo não temporal	54
5.3	Modelo híbrido.....	55
5.4	Análise de sensibilidade do modelo previsional	56
5.5	Cartas de controlo <i>short run</i> nas captações	58
5.6	Controlo em tempo real da qualidade da previsão do modelo	63
5.6.1	Fase I do controlo da previsão.....	67
5.6.2	Fase II do controlo da previsão	70
5.7	Capacidade do processo	75
Capítulo 6 – Conclusões e recomendações		77
6.1	Conclusões	77

6.2	Sugestões para trabalhos futuros	79
Bibliografia	81
Anexos	85
Anexo I – Tabelas estatísticas e de cartas de controlo		85
Anexo II – Tratamento de input e outputs.....		89
Anexo III – Cartas de controlo e modelo ARIMA		92

Índice de figuras

Figura 2.1 - Organograma SMAS Almada. Fonte: SMAS Almada, n.d.b	6
Figura 2.2 - Sistema de distribuição de água. Adaptado de: SMAS Almada, n.d.e	7
Figura 2.3 - Sonda de medição.....	9
Figura 3.1 - Sensor de 4 eléctrodos. Fonte: Ramos et al., 2008	12
Figura 3.2 - Função perda de qualidade. Adaptado de: Taguchi et al., 2004.	13
Figura 3.3 - Exemplo de um correlograma. Fonte: Brockwell & Davis, 2002	16
Figura 3.4- Exemplo de carta de média e amplitude. Fonte: Oakland, 2007	24
Figura 3.5 - As três zonas das cartas de controlo. Fonte: Oakland, 2007	24
Figura 3.6 - Limites de especificação. Fonte: Oakland, 2007	26
Figura 4.1 - Correção de uma série pelo <i>statistica</i>	44
Figura 4.2 - Evolução do caudal de uma entrada do sistema	44
Figura 4.3 - Fluxograma do modelo simplificado.....	46
Figura 4.4 - Fluxograma de uma bifurcação	47
Figura 4.5 - Esquema simplificado de uma rede de abastecimento	48
Figura 5.1 - Previsão da condutividade com dados reais	51
Figura 5.2 - Previsão e Valor real da condutividade num reservatório com inputs estimados ...	52
Figura 5.3 - Previsão e valor real da condutividade no reservatório com inputs reais a cada 33 minutos.....	53
Figura 5.4 - Comparação entre valor previsto e valor real da condutividade no modelo não-temporal.....	54
Figura 5.5 - Previsão e valor real da condutividade para o modelo híbrido com dados reais	55
Figura 5.6 - Previsão e valor real da condutividade para o modelo híbrido com dados estimados	56
Figura 5.7 - Condutividade da entrada E4.....	59
Figura 5.8 - Carta Q(X) da entrada E4.....	59
Figura 5.9 - Carta Q(MR) da entrada E4.....	60
Figura 5.10 - Carta W(MR) para a entrada E4	60
Figura 5.11 - Série de dados da condutividade da entrada E1.....	61
Figura 5.12 - Carta Q(X) do último patamar da entrada E1	62
Figura 5.13 - Carta Q(MR) do último patamar de valores da condutividade da entrada E1	62
Figura 5.14 - Série dos desvios da previsão a modelar	63
Figura 5.15 - Correlograma da série dos desvios da previsão	64
Figura 5.16- PACF da série dos desvios da previsão	64
Figura 5.17 - Série do termo do erro do modelo ARIMA (2; 1; 0)	65
Figura 5.18 - Ajustamento à distribuição normal de 100 observações do termo do erro do modelo ARIMA	66
Figura 5.19 - Primeira iteração da carta X para o controlo do termo erro do modelo ARIMA ..	67
Figura 5.20 - Primeira iteração da carta MR para o controlo do termo erro do modelo ARIMA	68
Figura 5.21 - Terceira iteração da carta X para o controlo do termo erro do modelo ARIMA...	69
Figura 5.22 - Terceira iteração da carta MR para o controlo do termo erro do modelo ARIMA	69
Figura 5.23 - Valor do somatório do erro ao quadrado em função de λ	71
Figura 5.24 - Carta MCEWMA para os valores previstos do modelo ARIMA	71
Figura 5.25 - Gráfico do índice T_s para os dados previstos do termo erro do modelo ARIMA..	72
Figura 5.26 - Carta MCEWMA com os dados observados do termo erro do modelo ARIMA ..	73
Figura 5.27 - Gráfico do índice T_s para os dados frescos da série do termo erro do modelo ARIMA	74

Figura 5.28 - Carta EWMA para os valores previstos do termo erro do modelo ARIMA.....	74
Figura 5.29 - Carta EWMA com os dados observados do termo erro do modelo ARIMA	75
Figura I.8.1 - Tabela qui-quadrado.....	85
Figura I.8.2 - Tabela Kolmogorov-Smirnov	86
Figura I.8.3 - Tabela Lilliefors	86

Índice de tabelas

Tabela 2.1 - Zonas de abastecimento por freguesias	8
Tabela 2.2 - Zonas de abastecimento e origens alternativas	8
Tabela 3.1 - Estimação dos parâmetros do modelo ARIMA. Adaptado de Box et al., 1994.....	21
Tabela 3.2 - Valores de k em função do tipo de especificação	33
Tabela 4.1 - Condutividades médias das captações subterrâneas do subsistema em estudo.....	49
Tabela 5.1 - Parâmetros do modelo temporal com os dados reais	52
Tabela 5.2 - Parâmetros do modelo temporal com os dados estimados.....	53
Tabela 5.3 - Parâmetros do modelo temporal com os dados reais com observações a cada 33 minutos.....	54
Tabela 5.4 - Parâmetros do modelo não-temporal com observações a cada 33 minutos	55
Tabela 5.5 - Parâmetros de erro do modelo híbrido com os dados reais.....	56
Tabela 5.6 - Parâmetros de erro do modelo híbrido com os dados estimados	56
Tabela 5.7 - Sensibilidade do modelo a variações na entrada E2	57
Tabela 5.8 - Sensibilidade do modelo a variações na entrada E3	57
Tabela 5.9 - Sensibilidade do modelo a variações na entrada E1	58
Tabela 5.10 - Parâmetros do modelo ARIMA da série dos desvios.....	65
Tabela 5.11 - Resultados obtidos do teste à normalidade pelo statistica.....	66
Tabela 5.12 - Parâmetros estimados para a série dos erros do modelo ARIMA	70
Tabela II.1 - Análise de sensibilidade da entrada E4	91

Capítulo 1 – Introdução

1.1 Enquadramento e motivação

A água potável é um bem natural que não existe em quantidade infinita no nosso planeta. Apesar do planeta Terra ser coberto por água, só 2,7% da mesma é doce. O ser humano consome mais de metade da água potável existente na terra, pois estima-se que existam, no nosso planeta, um total de 7000 m³ per capita, sendo que em 1950 esse valor era mais do dobro. A população humana e a sua procura por água potável não pararam de crescer desde então.

A água é uma necessidade crítica para toda a vida. Dessa forma a sua protecção e utilização de forma sustentável devem ser uma prioridade absoluta.

Neste contexto, a União Europeia, através da Carta Europeia da Água (“European Charter on Water Resources”, 2001) define os princípios sob os quais devem ser geridas a distribuição e consumo de água, nas suas diversas vertentes (consumo humano, industrial, etc.).

Nesse documento a União Europeia estipula que a água deve ser distribuída de forma equitativa e utilizada razoavelmente no interesse público, que as políticas e leis aplicáveis devem defender acima de tudo os ecossistemas aquáticos, que todos os seres têm direito a uma quantidade de água suficiente para cumprir as suas necessidades básicas, que os parceiros públicos ou privados devem ter práticas de gestão integradas de água, de modo a planear e distribuir a mesma de forma equitativa e sustentável, além de economicamente racional, entre outros princípios.

Nesse espírito, e em colaboração com os Serviços Municipalizados de Água e Saneamento do concelho de Almada (SMAS de Almada), nasce a presente dissertação, com o objectivo de ser um contributo ao controlo da qualidade da água potável, promovendo a gestão eficaz deste escasso e precioso recurso.

A qualidade da água de consumo é alvo do estudo do ser humano desde o século XIX, causado em grande parte pelos avanços científicos da época e pelo rápido crescimento dos centros urbanos. Nos Estados Unidos, os primeiros organismos criados com o objectivo de garantir a saúde pública através da monitorização da qualidade da água potável nas redes de distribuição conseguiram, entre outros feitos, determinar que a cólera, sendo uma doença transmitida pela água, podia ter os seus surtos reduzidos através de um eficaz controlo e melhoria dos métodos de distribuição da água potável (Edzwalds (ed. lit.), 2010, pp.1.2).

A crescente preocupação com o tratamento da água potável, isto é, o controlo na recolha, tratamento, filtração, desinfecção e distribuição é considerado um dos 10 grandes avanços na saúde pública do século XX.

A procura por água potável a nível mundial encontra-se em crescimento, as fontes de água potável são cada vez mais raras, o custo de tratamento da água cresce devido a novos contaminantes detectados e as leis que regulam os parâmetros de qualidade da água são cada vez mais exigentes. Neste cenário torna-se imperioso que as entidades que trabalham na área sejam capazes de lidar com estes desafios com novas técnicas de controlo da qualidade da água, de forma a minimizar o impacto destes factores nos seus custos de operação e suas consequências a nível de saúde pública.

1.2 Objectivos

A presente dissertação tem como objectivo a monitorização e previsão da qualidade da água, utilizando para esse efeito a sua característica de condutividade eléctrica, nos SMAS do concelho de Almada.

Pretende-se monitorizar, a partir de dados reais, os valores da característica da água recolhida em captações subterrâneas, utilizando para o efeito ferramentas estatísticas como as cartas de controlo.

Assentando em dados históricos, pretende-se construir um modelo que seja capaz de prever, em função dos dados das entradas (as captações ou reservatórios), as características de qualidade da saída da rede de tubagens do sistema de distribuição.

Uma variável que estará presente no modelo é o tempo. Pretende-se determinar como é que as diferentes configurações da rede (com comprimentos distintos) influenciam o resultado da condutividade da água no reservatório final. Sendo que as águas proveniente de fontes distintas têm características diferentes entre si, e cada uma demorará diferentes intervalos de tempo a chegar ao reservatório. O modelo deve ter essas discrepâncias em conta.

A monitorização das entradas de água no sistema e da eficácia da previsão para as características de saída devem ser realizadas através de cartas de controlo, tornando possível a existência de alarmes quando houver perturbações no processo. Estas perturbações podem ser contaminações nas captações ou roturas na rede de distribuição, que terão um efeito negativo na qualidade da água fornecida. Efeito esse que será visível no valor da condutividade da água.

A partir da monitorização das entradas e da previsão das saídas, pretende-se que seja possível controlar a qualidade da água para consumo humano ao longo do processo, desde a sua captação até à estação elevatória. Desta forma espera-se que seja possível detectar eventuais contaminações nas fontes ou nas condutas.

1.3 Estrutura da dissertação

A dissertação está dividida em 6 capítulos.

O capítulo I introduz o problema em estudo e refere os objectivos que se pretendem atingir com o modelo criado e sua aplicação nas redes do SMAS.

O Capítulo II introduz a entidade em estudo, como se encontra organizada, quais os seus objectivos e missão e demonstra a importância do presente estudo como contributo a atingir os seus objectivos.

O Capítulo III apresenta os conceitos teóricos, matemáticos e estatísticos necessários para a construção do modelo utilizado para a resolução do problema apresentado.

O Capítulo IV apresenta a metodologia seguida, em linha com os conceitos apresentados no capítulo III, para desenvolver os modelos de previsão que são utilizados neste estudo.

O capítulo V apresenta os resultados obtidos através da aplicação do modelo desenvolvido no capítulo IV, e respectiva discussão, analisando a qualidade dos resultados obtidos e sua aplicabilidade.

O capítulo VI apresenta as conclusões que podem ser retiradas das observações retiradas do capítulo V e apresenta sugestões para a sua aplicação nas redes dos SMAS de Almada, além de recomendações para trabalhos futuros.

Capítulo 2- SMAS de Almada

Neste capítulo pretende-se introduzir a entidade gestora na qual foi realizado o presente estudo, a forma como a mesma se organiza, a sua missão e gestão. Aborda-se ainda o seu funcionamento operacional, apresentando a sua rede de distribuição, na qual este estudo incide, e os seus meios tecnológicos de recolha de dados que serviram como base à realização da presente dissertação.

2.1 A entidade gestora

Os SMAS (Serviços Municipalizados de Água e Saneamento) nascem em 1950, por uma portaria que aprova a municipalização dos serviços de captação, distribuição e condução de água, e começam o seu funcionamento em Portugal em 1951, na altura com a designação de SMA (Serviços Municipalizados de Água). Só em 1957 a recolha de águas domésticas e pluviais é também municipalizada, sendo que os serviços tomam a designação que hoje ainda lhes é atribuída (SMAS Almada, n.d.-a).

Os SMAS são serviços públicos de interesse local com total autonomia em cada município de actuação (SMAS Almada, n.d.-b). As suas principais actividades são:

- Captação, adução, controlo de qualidade e distribuição de água potável;
- Recolha, drenagem e tratamento de águas residuais domésticas;
- Recolha e drenagem de águas pluviais;
- Exploração das redes de água (potável e/ou residual) e estruturas de apoio, com a sua respectiva construção, ampliação e manutenção.

A presente dissertação foi realizada em cooperação com o SMAS do município de Almada.

2.1.1 Estratégia da empresa/entidade

Os Serviços Municipalizados de Água e Saneamento do concelho de Almada tem uma linha estratégica que decorre do plano estratégico do município de continuar a desenvolver-se de forma mais sustentável e ecológica. Nesse seguimento, o SMAS de Almada tem como principais linhas de orientação estratégica:

- Prosseguir uma gestão transparente e pública, pautada pela solidariedade e eficiência;
- Diversificação constante das origens de água e preservação das já existentes;
- Cobrir integralmente o Concelho com o sistema de águas residuais e melhorar a capacidade de resposta do sistema de drenagem de águas pluviais;
- Optimizar o grau de operacionalidade do sistema de adução, armazenamento e distribuição de água;
- Garantir a qualidade da água fornecida.

É no seguimento dos dois últimos pontos da orientação estratégica que surge o estudo descrito na presente dissertação.

2.1.2 Gestão e Organograma da empresa/entidade

Os SMAS do município de Almada serviram, ao longo do ano de 2012, um total de 105925 clientes (entre clientes domésticos, comércio e indústria, associações ou colectividades e institucionais), aos quais facturou um total de 12.257.969 m³ de água (SMAS Almada, n.d.-c).

No início de 2013 os SMAS de Almada contavam com um total de 506 trabalhadores, onde mais de metade ocupam a função de Assistente Operacional (276 trabalhadores), seguidos pelos Assistentes Técnicos (102 trabalhadores). Existem ainda 66 Técnicos Superiores e 24 Especialistas e Técnicos Informáticos (SMAS Almada, 2013).

O organograma segundo o qual estão relacionais as secções funcionais na empresa pode-se encontrar na Figura 2.1.

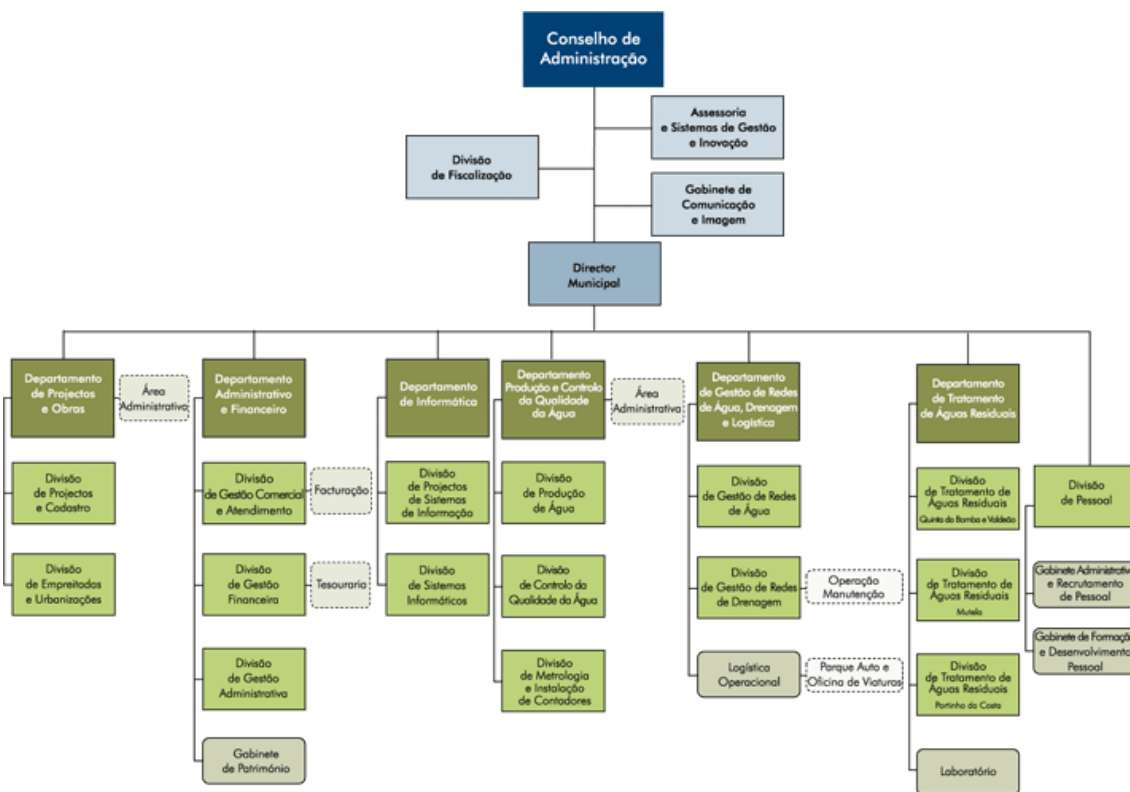


Figura 2.1 - Organograma SMAS Almada. Fonte: SMAS Almada, n.d.b

No orçamento previsional para 2013 os SMAS de Almada prevêem receitas correntes na ordem dos 25 milhões de euros e receitas de capital na ordem dos 6 milhões, que em conjunto com outras receitas expectáveis resultam num valor de receitas previstas na ordem dos 35,7 milhões de euros, sendo as despesas expectáveis um valor de ordem semelhante, em linha com a estratégia de gestão responsável e eficaz (SMAS Almada, n.d.-d).

2.2 Rede de abastecimento de água

Como já foi referido na secção 2.1, o SMAS de Almada é responsável pela captação, adução, reserva e distribuição da água.

A captação é efectuada com recurso a 33 furos subterrâneos de profundidade variável, sendo estes concentrados essencialmente no concelho do Seixal (localizados na Quinta da Bomba, Corroios e Vale Milhaços), sendo que esses furos fornecem o total de água consumida no Concelho.

A adução ocorre com o apoio de um sistema de condutas com um total de 84Km de comprimento e 9 estações elevatórias. A água captada nos furos subterrâneos é transportada até uma das estações elevatórias primárias (Vale de Milhaços, Corroios e Quinta da Bomba), de

onde são elevadas para um reservatório, através de estações elevatórias secundárias (que se encontram no Laranjeiro, Feijó, Lazarim, Cassapo, Raposo e Pragal). A partir desses reservatórios a água é distribuída de forma a satisfazer as necessidades dos consumidores do concelho de Almada.

Ao mesmo tempo existe uma reserva de água, de modo a que seja possível responder de forma mais eficaz a flutuações na procura e potenciais avarias ou necessidade de manutenções na rede de distribuição. Esta é constituída por 25 reservatórios, com uma capacidade total de aproximadamente 85000 m³/dia de água.

A distribuição de água é realizada através de uma rede de 15 sistemas de condutas com um total de 882 km de extensão, que asseguram a satisfação das necessidades de água directamente ao consumidor (SMAS Almada, n.d.-e).

Este sistema encontra-se resumido na Figura 2.2.

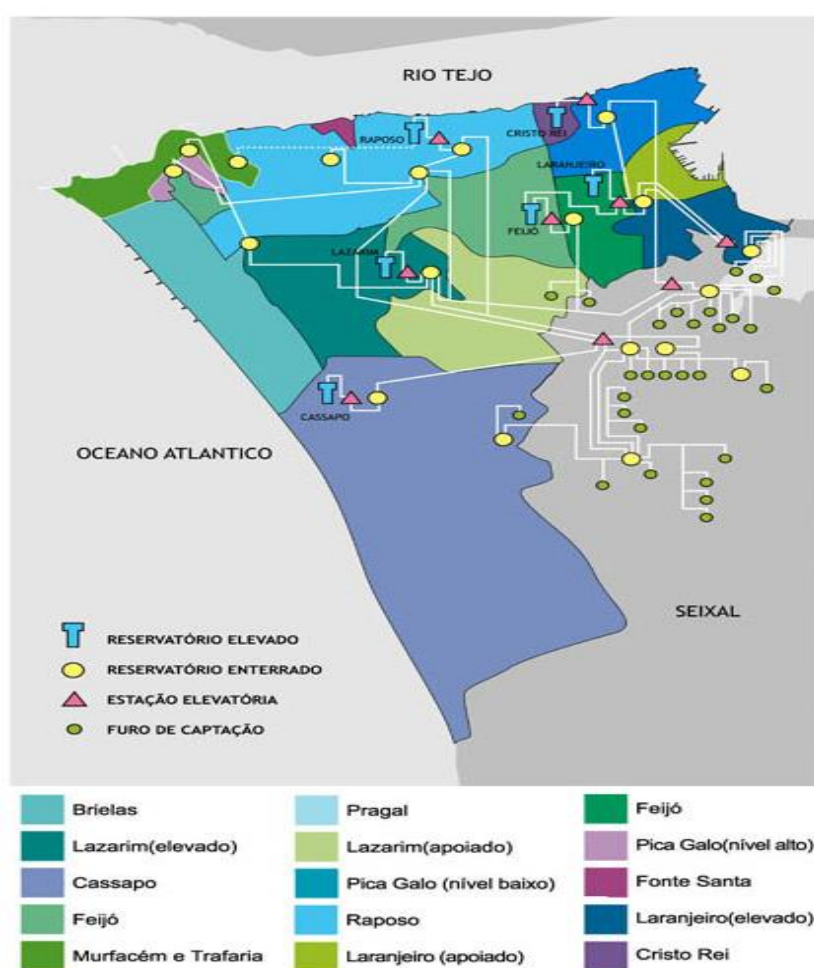


Figura 2.2 - Sistema de distribuição de água. Adaptado de: SMAS Almada, n.d.e

Esta rede de distribuição de água serve aproximadamente 174000 habitantes, captando um total anual de 16929000 m³ de água, com uma capacidade de produção diária de 77880 m³ de água.

Os SMAS de Almada distinguem um total de cinco zonas de abastecimento, resumidas na Tabela 2.1.

Tabela 2.1 - Zonas de abastecimento por freguesias

Zona de Abastecimento	Freguesias
ZA1	Almada, Pragal, Cacilhas, Cova da Piedade (Nordeste)
ZA2	Laranjeiro, Cova da Piedade e Feijó (Norte)
ZA3	Feijó e Laranjeiro
ZA4	Costa da Caparica, Sobreda, Trafaria e Caparica
ZA5	Charneca da Caparica e Costa da Caparica

Para efeitos de reserva, existem interligações entre algumas zonas de abastecimento, de modo a melhor satisfazer variações na procura. As interligações encontram-se descritas na Tabela 2.2.

Tabela 2.2 - Zonas de abastecimento e origens alternativas

Zona de Abastecimento	Origens alternativas
ZA1	ZA2 e ZA4
ZA2	ZA1 e ZA3
ZA3	ZA2 e ZA4
ZA4	ZA1, ZA3 e ZA5
ZA5	ZA4

2.3 Telegestão

Os SMAS de Almada estão actualmente a terminar a 3ª fase do processo de implementação de um sistema informático que permite em tempo real e através da internet o controlo das variáveis e parâmetros relativos à exploração de furos de captação, dos seus reservatórios e estações elevatórias de forma integrada (SMAS Almada, n.d.-f).

Este sistema permitirá controlar ainda os níveis de água nos reservatórios e aquíferos, estados de funcionamento de electrobombas, entre outros parâmetros operacionais, constituindo ainda um método automático de prevenção de acidentes, pois tem capacidade de detecção de fugas de cloro, tomando as medidas necessárias para garantir as condições de segurança, tal como sistemas anti-intrusão. Permite ainda detectar roturas na rede distribuidora, minimizando perdas resultantes da fuga de água.

No âmbito deste estudo, a telegestão terá um papel fulcral na recolha de dados dos parâmetros da qualidade da água na rede distribuidora, tal como medidores de caudal, de forma a que seja possível caracterizar o estado da rede e dos reservatórios em tempo real.

2.4 Controlo da qualidade da água nos SMAS de Almada

Os SMAS de Almada encontram-se equipados com um laboratório com capacidades de análise de água para consumo humano e água de piscinas.

Este laboratório efectua anualmente cerca de 17000 análises de controlo da qualidade da água, entre análises de rotina obrigatórias e de controlo operacional.

O laboratório tem implementado um Sistema de Garantia de Qualidade de acordo com os requisitos da norma NP EN ISO 9001:2000, sendo que prepara ainda a candidatura à acreditação da qualidade pela norma NP EN ISO 17025:2005, norma que estabelece os requisitos gerais de competência para laboratórios de ensaio e de calibração.

Este laboratório participa ainda em ensaios interlaboratoriais nacionais e europeus nas áreas da química e microbiologia de águas de consumo (SMAS Almada, n.d.-g).

Os SMAS de Almada seguem um plano por si elaborado, o Programa de Controlo da Qualidade da Água, com colheitas regulares em mais de 450 pontos estratégicos no consumidor final, sendo que os resultados das análises das amostras são submetidos à Entidade Reguladora de Serviços de Águas e Resíduos, sendo que esta entidade é responsável pela aprovação anual desses mesmos resultados (*Decreto-Lei n.º306, 2007*). Ao mesmo tempo efectua um controlo operacional nas captações, nas centrais de abastecimento e elevação e noutros pontos da rede (SMAS Almada, n.d.-h).

Cada uma das zonas de abastecimento é alvo de análises de controlo de rotina e de inspecção, cujo objectivo é controlar parâmetros físicos, químicos e microbiológicos, cuja frequência mínima se encontra definida em decreto-lei (*Decreto-Lei n.º236, 1998*).

Na Figura 2.3 mostra-se o equipamento utilizado pelo SMAS de Almada para medir, além de outras variáveis, a condutividade e o caudal da água que passa numa determinada conduta, desviando da mesma um pequeno volume de água. Esse pequeno volume é imediatamente devolvido à conduta, sendo que os valores medidos pela sonda são enviados em tempo real para a telegestão.



Figura 2.3 - Sonda de medição

Capítulo 3 – Fundamentos Teóricos

Neste capítulo pretende-se abordar os conhecimentos indispensáveis à realização desta dissertação, que serviram como base à construção dos modelos implementados nos capítulos seguintes.

3.1 A Condutividade da água

A condutividade de um material é definida como a capacidade desse mesmo material permitir a passagem de corrente eléctrica (US EPA, n.d.). A condutividade eléctrica da água (ou, no contexto da presente dissertação, apenas condutividade) é uma característica física da mesma que pode ser utilizada como indicador da presença de substâncias inorgânicas como cloro, nitratos e outros sais e minerais. Quanto maior a condutividade específica de uma amostra de água, maior será a concentração do total de sais nela dissolvidos (Edzwalds (ed. lit.), 2010, p. 3.3). Essa concentração será um parâmetro que terá implicações na qualidade da água, e como tal, deve ser controlado.

Alterações significativas a um valor de condutividade de uma fonte de água poderão indicar a presença de fontes de poluição indesejadas, dado que uma fonte de água sem perturbação terá um valor de condutividade aproximadamente constante, a uma temperatura constante, uma vez que a condutividade da água é também função da temperatura da mesma.

Dessa forma o valor da condutividade de uma amostra de água torna-se num indicador da sua qualidade utilizado largamente na indústria com diversas aplicações práticas, como por exemplo análise de fugas de água e análise de águas de irrigação. O método da estimação da concentração de sais presentes numa amostra de água em função do seu valor da condutividade específica é considerado como sendo um método que resulta numa boa aproximação do seu valor real e mais eficaz que outros métodos existentes, pois os seus resultados são influenciados por todos os sais potencialmente presentes na amostra. Outros métodos, como por exemplo o método que consiste em colocar em ebulição um volume definido da amostra de modo a analisar o resíduo deixado, resultam em que a presença de certos nitratos passe despercebida (Gustafson & Behrman, 1939).

Para medir a condutividade da água é necessário uma amostra a uma temperatura definida (20 °C ou 25 °C) (Ramos *et al.* 2008). Essa amostra (que se pressupõe uma solução) é submetida a um sensor que irá medir a condutividade.

Existem dois tipos de sensores de condutividade: toroidais ou com eléctrodos. Os sensores de eléctrodos podem ter dois, três ou quatro eléctrodos. A condutividade, que é inversa à resistência eléctrica será função da corrente eléctrica que atravessa a solução a partir dos eléctrodos e de uma constante de proporcionalidade que depende do número e da geometria dos eléctrodos. Um exemplo de um sensor com 4 eléctrodos pode ser encontrado na Figura 3.1.



Figura 3.1 - Sensor de 4 eléctrodos. Fonte: Ramos et al., 2008

Os sensores toroidais contém normalmente duas bobinas envoltas numa estrutura não condutora, onde uma bobina induz uma corrente eléctrica na água e a segunda mede a corrente induzida, que é proporcional à condutividade da solução em estudo.

A unidade do sistema internacional de unidades (SI) para a condutividade é o Siemens por metro (S/m), mas a unidade utilizada neste estudo será microSiemens por centímetro ($\mu\text{S}/\text{cm}$).

A legislação portuguesa aponta como parâmetro indicador da condutividade da água para consumo humano, ou seja, o valor que deve ser tomado como guia, o de $2500 \mu\text{S}/\text{cm}$ a 20°C (*Decreto-Lei n.º 306, 2007*).

3.2 A Qualidade

Não existe uma única definição de qualidade, pois é um conceito demasiado abrangente para tal. Existem diversas definições que se encaixam dependendo do contexto em que se analisa a qualidade de um produto, pois esta contém em si demasiadas características díspares, que se podem encaixar, de grosso modo, em dois grandes grupos (Pires, 2007, p.21):

- Características funcionais;
- Características técnicas.

As características funcionais são definidas como sendo aquelas que o consumidor estará directamente interessado, como a comodidade, fiabilidade, entre outras. Por outro lado, as características técnicas compreendem as que resultam de uma opção técnica tomada para concretizar o produto, como material, peso, altura, entre outras.

Não se pode considerar que exista a qualidade absoluta. A qualidade de um produto será sempre subjectiva, pois é função do mercado a que se destina, deve ter capacidade de satisfazer os seus consumidores, tendo as características para prestar o serviço ou as tarefas que os mesmos esperam dele ao custo que os mesmos estão dispostos a pagar.

No estudo da qualidade, autores conceituados definem qualidade de formas bastante distintas. Juran (Juran & Godfrey, 1998) define qualidade como dois conceitos, um deles é o produto/serviço cumprir especificações (dessa forma seguindo as necessidades do consumidor, consequentemente satisfazendo-o) e o outro é o produto/serviço estar livre de defeitos. O autor sugere ainda três passos para a gestão para a qualidade, criados tendo como base a gestão financeira:

- Planeamento da qualidade;
- Controlo da qualidade;
- Melhoria.

Por outro lado, Taguchi (2004, p.171) propõe que a qualidade seja definida como uma função de perda, a que chama função de perda de qualidade. Essa perda tanto pode ser para o produtor como para o consumidor (sociedade). O autor considera que a qualidade de um produto/serviço é máxima quando a perda é mínima, ou seja, o valor de uma característica de qualidade é igual ao seu valor alvo (*target*), exemplificado na Figura 3.2.



Figura 3.2 - Função perda de qualidade. Adaptado de: Taguchi et al., 2004.

Deming, outro autor histórico da área da qualidade, citado por Pires (2007, p.27) propõe uma abordagem onde as necessidades dos consumidores serão o ponto de partida para a melhoria da qualidade, dando origem à ideia de melhoria contínua, e apresenta um modelo que fica conhecido pelos 14 pontos de Deming, nos quais se incluem:

- Fim da dependência da inspecção para garantia de qualidade;
- Redução de variação, trabalhando com o mínimo de fornecedores;
- Formação no posto de trabalho;
- Eliminação de barreiras funcionais entre áreas distintas das organizações.

Pires (2007, p.27) cita ainda Crosby, que considera que a qualidade é «um problema de pessoas», onde identifica 6 factores-chave para a gestão da qualidade:

1. Qualidade não deve significar elegância mas sim conformidade;
2. Não existem problemas de qualidade;
3. É mais económico fazer bem à primeira tentativa;
4. A medida de desempenho da qualidade será os custos com a mesma;
5. O objectivo será zero defeitos;
6. Qualidade não tem custos.

Pires (2007, p.27) considera que, independentemente de qual a filosofia seguida, as abordagens são consistentes no que toca à necessidade de melhoria contínua, que a gestão para a qualidade é constituída por um conjunto de atitudes onde a gestão de topo tem responsabilidade máxima e que a orientação para a gestão da qualidade de uma qualquer organização é algo que deve ser gerida num prazo alargado.

3.3 As sete ferramentas básicas da qualidade

Sendo que as questões sobre a qualidade podem vir de um leque tão díspar de temas, cada qual com as suas especificações, há diversas formas de analisar problemas de qualidade. Ainda assim, para a gestão da qualidade, considera-se que existem sete ferramentas básicas, as chamadas sete ferramentas da qualidade, com as quais se poderá abordar todo o tipo de situações no que toca à gestão da qualidade (Pereira & Requeijo, 2012, p.34).

As sete ferramentas da qualidade são:

1. Fluxograma;
2. Folhas de registo e verificação;
3. Histograma;
4. Diagrama de Pareto;
5. Diagrama de Causa-e-Efeito (ou diagrama de Ishikawa);
6. Gráfico de dispersão
7. Cartas de Controlo (Controlo estatístico do processo).

O presente estudo foca-se na utilização da sétima ferramenta, o controlo estatístico do processo, utilizando cartas de controlo, para controlar a água recolhida das captações subterrâneas, e os valores da previsão das características da mesma nos reservatórios da rede dos SMAS de Almada.

3.4 Análise de séries temporais

A presente dissertação foca-se essencialmente em dados sob a forma de séries temporais. Uma série temporal é definida como sendo uma sequência de dados recolhidos de forma sequencial no tempo (Box et al., 1994, p.1). Inúmeros tipos de dados podem surgir sob a forma de séries temporais, em campos tão distintos como a economia, engenharia, ciências naturais, meteorologia, ciências sociais, entre outros.

Um factor intrínseco a uma série temporal é que, tipicamente, observações adjacentes são dependentes, e a natureza dessa dependência tem muito interesse prático.

Uma das grandes motivações para o estudo das séries temporais é a possibilidade de, a partir de dados do passado e actuais se poder prever o comportamento da série no futuro.

As séries temporais podem-se considerar como sendo de dois tipos distintos:

- Série estacionária;
- Série não estacionária.

Uma série estacionária é definida como sendo uma série cujos dados apresentam um comportamento que segue uma média aproximadamente constante, ou seja, o processo que descreve está em equilíbrio (Box et al., 1994, p. 7). Com a crescente atenção da indústria a este tema, surgem outro tipo de séries temporais com comportamentos que não seguem qualquer distribuição aparente ao longo do tempo. Este tipo de séries são conhecidas como séries não estacionárias, e merecem uma análise estatística mais cuidada. Um dos modelos utilizados para trabalhar com séries não estacionárias são os modelos ARIMA, popularizados por Box e Jenkins em 1970 (Makridakis et al., 1997, p. 336).

Em análise de dados, especialmente quando os mesmos se apresentam sob a forma de série temporal, é comum encontrar-se uma situação em que a evolução de uma dada variável está relacionada com a evolução de uma outra variável. Quando esta situação ocorre, diz-se que as variáveis estão correlacionadas (Makridakis et al., 1997, p. 37). O facto de se considerar duas variáveis correlacionadas não deve implicar, no entanto, que haja uma relação de dependência entre as mesmas.

3.4.1 Correlação de dados

De modo a que seja possível quantificar a relação entre duas variáveis, define-se uma variável que se chama o coeficiente de correlação. O valor deste coeficiente varia entre 1 e -1, sendo que ao valor 1 corresponde uma correlação perfeita positiva (quando uma variável aumenta, a outra também aumenta), o valor -1 corresponde a uma correlação perfeita negativa (quando uma variável aumenta, a outra diminui) e ao valor 0 corresponde a falta de correlação, ou seja, as variáveis são completamente independentes. O aumento (em módulo) do coeficiente de correlação indica a força da associação entre as duas variáveis. É necessário ainda referir que o coeficiente de correlação apenas mede a associação linear entre as variáveis. É possível haver casos de variáveis com coeficiente de correlação aproximadamente nulo mas entre as quais exista uma relação não linear.

O coeficiente de correlação entre duas variáveis X e Y é dado pela equação (3.1) .

$$r_{XY} = \frac{COV(X,Y)}{S_X S_Y} \quad (3.1)$$

$$COV(X,Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (3.2)$$

onde:

r_{XY} – Coeficiente de correlação entre X e Y;

S_X – Desvio-padrão da variável X;

S_Y – Desvio-padrão da variável Y;

$COV(X,Y)$ – Covariância entre as variáveis X e Y, dada pela expressão (3.2).

n – Tamanho da amostra em estudo;

De notar que a presença de uma observação “anormal” (cuja definição será dada no subcapítulo 3.11) pode resultar no que se chama de efeito “King Kong”, que ocorre quando uma observação de valor significativamente diferente do restante conjunto de dados tem um grande efeito no valor do coeficiente de correlação.

3.4.2 Autocorrelação de dados

Enquanto que o coeficiente de correlação mede uma relação explicativa entre duas variáveis, existe um coeficiente que exprime semelhante relação para uma só variável e si própria, mas atrasada de um determinado número de períodos.

Comparando, por exemplo, uma variável num determinado instante no tempo e no instante de tempo anterior é possível observar uma eventual relação entre observações consecutivas. À observação no período anterior chama-se de observação com desfasamento 1 ou lag 1. Seguindo a mesma lógica, é possível comparar observações com lag 2, 3, etc.

Dessa forma define-se o coeficiente de autocorrelação para um dado valor de lag pela expressão da equação (3.3).

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=k+1}^n (Y_t - \bar{Y})^2} \quad (3.3)$$

Onde:

r_k - Coeficiente de autocorrelação para lag k;

Y_t - Observação da variável Y no instante t;

Y_{t-k} - Observação da variável Y no instante t-k;

\bar{Y} - Valor médio de Y.

Ao conjunto dos valores do coeficiente de autocorrelação para valores de desfasamento de 0, 1, 2, ... designa-se por Função de Autocorrelação (*Autocorrelation function* ou ACF na terminologia inglesa). Construindo o gráfico da evolução da função de autocorrelação em função do desfasamento k obtem-se o diagrama conhecido como correlograma (Pereira & Requeijo, 2012, p.643). Um exemplo de um correlograma pode-se encontrar na Figura 3.3. Deste exemplo de correlograma poder-se-ia concluir que existe uma forte correlação entre as variáveis apenas para uma *lag* de uma unidade de tempo.

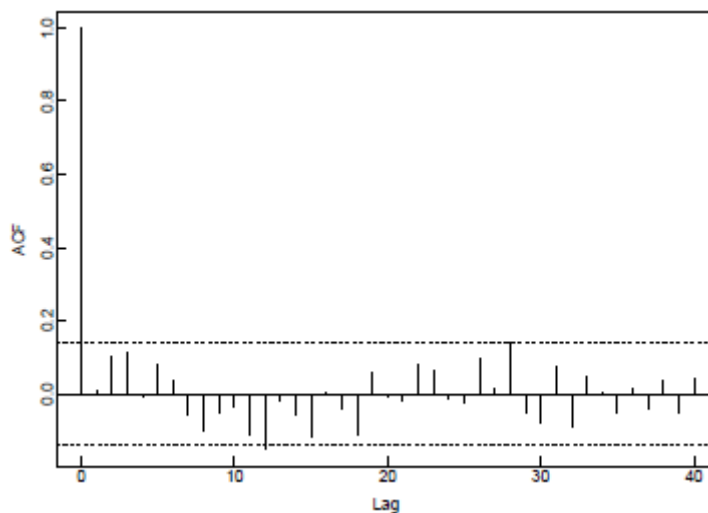


Figura 3.3 - Exemplo de um correlograma. Fonte: Brockwell & Davis, 2002

Considera-se que se os valores da ACF pertencerem todos a um determinado intervalo, pode-se considerar que a série de dados é independente, ou seja, que os dados não estão autocorrelacionados. A definição desse intervalo de valores não é um assunto consensual entre autores da temática. Pereira & Requeijo (2012, p.643) defendem que o intervalo de valores deve ser um intervalo de confiança que é função da variância do coeficiente de autocorrelação, do tamanho da amostra e de um dado nível de significância α , dado pela expressão (3.4), enquanto

que Brockwell (2002, p.20) e Makridakis (1997, p.318) defendem que os mesmos devem ser apenas função do tamanho da amostra. A expressão (3.6) é a referida pelos dois últimos autores. Outra forma mais simples de considerar que existe autocorrelação será se o valor do coeficiente de autocorrelação for superior a um nível de significância α , normalmente 0,5. Este é o método mais utilizado pela maioria dos *softwares* de estatística.

$$P\left(-Z_{\frac{\alpha}{2}}\sqrt{\hat{v}} \leq r_k \leq Z_{\frac{\alpha}{2}}\sqrt{\hat{v}}\right) = 1 - \alpha \quad (3.4)$$

$$\hat{v} = Var(r_k) \approx \frac{1}{n} \left(1 + 2 \sum_{i=1}^{k-1} r_i^2 \right) \quad (3.5)$$

$$P\left(-\frac{1,96}{\sqrt{n}} \leq r_k \leq \frac{1,96}{\sqrt{n}}\right) = 1 - \alpha \quad (3.6)$$

onde:

\hat{v} – Estimador da variância de r_k , cujo valor é dado pela expressão (3.5);

$Z_{\frac{\alpha}{2}}$ – Distribuição normal reduzida com um nível de significância de $\alpha/2$;

n – Tamanho da amostra.

De notar que estes limites só permitem concluir se para um determinado desfasamento existe ou não autocorrelação significativa (Makridakis et al., 1997, p.319). De forma a determinar se no conjunto de dados como um todo existe autocorrelação, pode-se utilizar um teste estatístico baseado na estatística Q de Ljung-Box, dado pela expressão da equação (3.7).

$$Q = n(n+2) \sum_{k=1}^h (n-k)^{-1} r_k^2 \quad (3.7)$$

onde:

h - Valor máximo de desfasamento a considerar. Makridakis (1997, p.318) menciona como indicador um valor de h igual a 20.

O valor da estatística Q de Ljung-Box é seguidamente comparado com o valor da distribuição Qui-Quadrado com h graus de liberdade, cuja tabela se encontra no anexo I. Se o valor desta estatística Q for superior ao valor crítico que pode ser retirado da tabela da distribuição Qui-Quadrado com h graus de liberdade e um dado nível de significância α , considera-se que a distribuição dos dados em estudo não é independente, ou seja, existe autocorrelação na série de dados. Caso se observe que o valor da estatística Q de Ljung-Box é inferior ao valor crítico, considera-se que a série de dados é independente, e a sua variabilidade é devida ao “ruído branco” (ou *white noise* na terminologia inglesa), ou seja, não existe qualquer padrão na série de dados.

3.4.3 Autocorrelação parcial

Na regressão linear múltipla existe o conceito de correlação parcial, isto é, quando uma variável Y é função de duas ou mais variáveis explicativas (X_1, X_2, \dots, X_n), é possível determinar o quanto uma determinada variável X_i explica sobre a evolução da variável Y sendo que os efeitos das restantes variáveis explicativas são retirados. A mesma lógica se aplica no caso de uma série temporal.

A autocorrelação parcial mede o grau de associação entre Y_t e um dos seus desfasamentos Y_{t-k} , sendo removidos os efeitos dos restantes desfasamentos possíveis.

Um dos métodos possíveis para determinar o coeficiente de autocorrelação parcial de ordem k , ou α_k , pode ser estimado através do coeficiente de regressão ϕ_k , que se encontra na equação (3.8).

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_k Y_{t-k} + e_t \quad (3.8)$$

onde:

ϕ_j – j -ésimo coeficiente de regressão;

Y_t – Valor da variável Y no instante t ;

e_t – erro da regressão ou resíduo, variável que é idêntica e igualmente distribuída.

A este tipo de regressão, ou seja, a regressão de uma variável em função das suas observações anteriores, chama-se de autoregressão (AR, ou modelo autoregressivo), e este conceito tem importância nos modelos utilizados neste estudo. De modo análogo ao coeficiente de autocorrelação, é possível definir uma função de autocorrelação parcial, e da mesma construir um gráfico semelhante ao correlograma (Makridakis et al., 1997, p.321).

3.4.4 Modelos ARIMA de Box-Jenkins

Diversas séries temporais de dados empíricos (como por exemplo o comportamento de um conjunto de ações) comportam-se como séries não estacionárias, necessitando de modelos estatísticos mais complexos para a sua análise. Um desses é o modelo ARIMA, ou *autoregressive integrated moving average*.

Este modelo é constituído por três componentes, sendo que uma delas já foi referida na secção 3.4.3, a autoregressão (AR). Os outros dois são *moving average* (MA) e a componente da integração (I). De notar que neste contexto, a terminologia *moving average* (ou média móvel) não deve ser confundida com a média móvel de observações.

Existem diversos modelos ARIMA, geralmente conhecidos como modelos ARIMA (p;d;q), em que:

- p – Ordem da componente autoregressiva;
- d – Ordem do diferencial envolvido;
- q – Ordem da componente *moving average*;

Seguindo a mesma lógica para a componente AR explicada na secção 3.4.3, também é possível definir uma regressão para a variável Y tendo como variáveis explicativas não os valores

passados de Y, mas os erros (resíduos) dos mesmos. Esta situação pode-se escrever sob a forma de uma expressão matemática, que se encontra na equação (3.9).

$$Y_t = \theta_0 + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} + e_t \quad (3.9)$$

onde:

θ_j – j-ésimo coeficiente do modelo MA(q);

e_i – erro da regressão ou resíduo no instante i.

Ao modelo descrito pela equação (3.9) chama-se *moving average* (MA).

Aos modelos compostos simultaneamente por uma componente AR e MA, chamam-se modelos ARMA, e os mesmos são capazes de descrever eficazmente comportamentos de séries estacionárias, mas não são suficientes quando a série em estudo é não estacionária.

Para adaptar estes modelos para a sua utilização em séries não estacionárias, alguns conceitos devem ser introduzidos, como o operador de desfasamento e o operador diferencial, descritos nas equações (3.10) e (3.11), respectivamente.

$$B^m Y_t = Y_{t-m} \quad (3.10)$$

$$\nabla Y_t = Y'_t = Y_t - Y_{t-1} = (1 - B)Y_t \quad (3.11)$$

onde:

B^m – Operador de desfasamento de m unidades de tempo;

∇Y_t ou Y'_t – Operador diferencial.

A aplicação do operador diferencial permite retirar à série temporal a sua característica não estacionária (Makridakis et al., 1997, p. 326). Um diferencial de segunda ordem pode ser escrito como $(1-B)^2 Y_t$ ou ainda como Y'' , e utilizado em situações onde a aplicação do operador diferencial uma vez não é suficiente para retirar à série a sua característica não estacionária.

Um modelo ARMA de ordem (p;q) pode-se encontrar sintetizado na expressão (3.12) (Makridakis et al., 1997, p. 342).

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (3.12)$$

onde:

c – Termo constante;

ϕ_j – j-ésimo parâmetro de autoregressão;

θ_j – j-ésimo parâmetro da média móvel.

Este modelo pode ser simplificado da forma descrita na expressão (3.15), utilizando para tal o operador de desfasamento e as simplificações das equações (3.13) e (3.14) (Box et al., 1994, p. 91).

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (3.13)$$

$$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q) e_t \quad (3.14)$$

$$\phi(B) Y_t = \theta(B) e_t \quad (3.15)$$

De modo a poder modelar séries não estacionárias, é necessário incluir a componente I no modelo, e isso pode ser conseguido através da aplicação do operador diferencial (Box et al., 1994, p. 93). Dessa forma, o modelo ARIMA de ordem (p;d;q) resulta na expressão (3.16).

$$\phi(B) \nabla^d Y_t = \theta(B) e_t \quad (3.16)$$

No respeitante à notação a utilizar na aplicação deste modelo, convencionou-se que um modelo ARIMA sem uma determinada componente (seja ela AR, I ou MA), a mesma não surge no nome do modelo. Quer isto dizer, que um modelo ARIMA onde não exista componente I pode-se denominar modelo ARIMA (p;0;q) ou ARMA(p;q). A mesma lógica se aplica para um modelo onde não exista componente autoregressiva, sendo denominado ARIMA (0;d;q) ou IMA(d;q).

Um modelo de ruído branco, ou seja, um modelo onde a observação num determinado momento seja independente de observações anteriores, não existe diferenciação e não existe componente *moving average* pode ser denominado como modelo ARIMA (0;0;0).

3.4.5 Estimação dos parâmetros p, d, q do modelo ARIMA

A dificuldade inicial para aplicação dos modelos ARIMA é a necessidade de definir os parâmetros de ordem da componente autoregressiva e *moving average*. Enquanto que é relativamente simples definir a ordem do componente do diferencial, pois basta que a série se torne estacionária para se encontrar a ordem do mesmo, os restantes dois parâmetros requerem análise das funções autocorrelação e autocorrelação parcial (ACF e PACF).

Através da evolução ACF e da PACF ao longo dos diversos valores de lag é possível inferir, com um determinado grau de confiança, os valores dos parâmetros componente autoregressiva e *moving average* (p e q, respectivamente), e qual dos modelos se deve incluir na análise, senão mesmo ambos. Box (1994, p. 84) apresenta uma tabela, que se resume na tabela Tabela 3.1, que analisa as características da ACF e PACF e aponta o modelo que mais se adequa.

Tabela 3.1 - Estimação dos parâmetros do modelo ARIMA. Adaptado de Box et al., 1994

	ACF	PACF
AR(p)	Decresce exponencialmente ou sinusoidalmente, de forma amortecida, a tender para zero;	Cai abaixo dos limites após a lag p;
MA(q)	Cai abaixo dos limites após a lag q;	Decresce exponencialmente ou sinusoidalmente, de forma amortecida, a tender para zero;
ARMA(p;q)	Decresce exponencialmente ou sinusoidalmente, de forma amortecida a partir da lag (q-p), a tender para zero;	Decresce exponencialmente ou sinusoidalmente, de forma amortecida a partir da lag (p-q), a tender para zero;

Um factor importante a ter em conta é que, na prática, raramente é necessário considerar valores de p, d ou q que não sejam 0, 1 ou 2, logo a identificação do modelo é simplificada imediatamente (Makridakis et al., 1997, p. 345).

Caso se verifique que não existe estacionariedade na ACF e na PACF, a série deve ser diferenciada o número de vezes necessária de modo a que se atinja a estacionariedade, e o número de vezes que se diferenciou a série será o valor do parâmetro d. Caso não seja claro se se deve aplicar um modelo MA ou AR, deve-se utilizar um modelo misto ARMA.

Outro factor importante é o número mínimo de observações a ter em conta para identificar correctamente o modelo que melhor se ajusta ao processo em estudo. O número mínimo de observações desejável será 100 (Pereira & Requeijo, 2012, p. 662). Mas deve-se respeitar sempre o princípio da parcimónia. A partir de um determinado momento, os custos financeiros e de tempo na recolha de mais amostras não trarão benefícios significativos para o modelo.

3.4.6 Média e variância de processos ARIMA

No caso de se estar perante um modelo AR(p), ou seja, um modelo auto-regressivo de ordem p, cuja expressão se encontra na equação (3.8), o valor esperado e a variância do processo (Y) a ser modelado são dados pela expressão (3.21) e (3.22), respectivamente.

$$E(Y) = \mu_Y = \frac{\phi_0}{1 - \sum_{j=1}^p \phi_j} \quad (3.17)$$

$$VAR(Y) = \sigma_Y^2 = \sigma_e^2 \frac{1}{1 - \sum_{j=1}^p \phi_j r_j} \quad (3.18)$$

onde:

σ_e – Desvio padrão do erro do modelo AR(p);

r_j – Coeficiente de autocorrelação para um desfasamento j .

Caso se esteja a estudar um modelo para o qual uma modelação $MA(q)$ seja mais adequada (média móvel de ordem q), cuja expressão se encontra na equação (3.9), o valor esperado e a variância do processo são dados pela expressão (3.19) e (3.20), respectivamente.

$$E(Y) = \mu_Y = \theta_0 \quad (3.19)$$

$$VAR(Y) = \sigma_Y^2 = \sigma_e^2 \sum_{j=1}^q \theta_j^2 \quad (3.20)$$

3.4.7 Teste à normalidade de uma série de dados

A distribuição normal é um tipo de distribuição de especial importância na gestão da qualidade, pois é uma condição *sine qua non* para a aplicação de diversas técnicas de controlo da qualidade, algumas das quais aplicadas neste estudo (Pereira & Requeijo, 2012, p. 135).

Como tal, apresentam-se duas formas de avaliar, através de um teste de hipóteses, se um determinado conjunto de dados segue ou não uma distribuição normal. Os dois testes aqui apresentados são o teste do Qui-Quadrado e o teste de Kolmogorov-Smirnov (ou teste KS). A maioria dos *softwares* de análise estatística têm estes testes incorporados, e permitem um teste mais eficaz da normalidade de uma população.

Sendo que ambos os testes são, na sua essência, testes de hipóteses, a primeira fase será definir a hipótese nula (H_0) e a hipótese alternativa (H_1). São elas:

H_0 : A população segue uma distribuição normal;

H_1 : A população não segue uma distribuição normal.

O teste do Qui-Quadrado, baseado na distribuição estatística que lhe dá o nome, permite a comparação da amostra em estudo com uma distribuição teórica. Consiste na comparação das frequências observadas e das expectáveis da distribuição teórica, e devolve uma estatística de teste que deve ser comparada com um valor tabelado da distribuição Qui-Quadrado, para um determinado valor de significância (α) e para um determinado número de graus de liberdade (v).

No teste do qui-quadrado deve-se rejeitar a hipótese nula quando o valor observado da estatística de teste for superior ao valor tabelado da distribuição qui-quadrado, para um determinado α e v . Caso se rejeite a hipótese nula, a conclusão a retirar do teste será que, para um nível de confiança $(1 - \alpha)$ a população em estudo não segue a distribuição normal.

O teste de Kolmogorov-Smirnov (ou teste KS) estuda o ajustamento da função de distribuição amostral à função de distribuição normal. Este teste é, geralmente, mais potente que o teste do Qui-Quadrado, mas é também mais complexo de calcular, embora a capacidade de cálculo dos *softwares* de estatística tenham contribuído para a eliminação dessa desvantagem. Por outro lado, o teste de Kolmogorov-Smirnov só é aplicável em dados contínuos, ao passo que o teste do Qui-Quadrado pode ser aplicado quer a dados discretos quer a dados contínuos.

Neste teste, à semelhança do teste do Qui-Quadrado, deve-se comparar o valor crítico com o valor da estatística de teste, para um determinado nível de significância α . O valor crítico pode ser obtido através de tabelas construídas para o efeito (no caso do teste do qui-quadrado, as suas tabelas encontram-se presentes no anexo I), sendo apenas necessário saber se os parâmetros da população são conhecidos à priori ou estimados a partir da amostra, pois isso terá influência na tabela a usar. Na primeira situação deve-se utilizar as tabelas KS, e no segundo caso, deve-se utilizar as tabelas criadas por Lilliefors. Ambas as tabelas se encontram presentes no anexo I.

No teste KS, deve-se rejeitar a hipótese nula caso o valor da estatística de teste observado seja superior ao valor retirado da tabela. Nessa situação, a conclusão a retirar será que, para um nível de confiança $(1 - \alpha)$ a amostra em estudo não segue a distribuição normal.

O valor da estatística de teste, tanto do teste do Qui-Quadrado como do teste KS pode ser obtido através de *softwares* como o *statistica*, tornando o teste da normalidade de uma amostra um processo simples e rápido de realizar.

3.5 Controlo Estatístico do Processo

O controlo estatístico do processo (ou SPC - *Statistical Process Control* na sigla inglesa) surge da necessidade de estabilizar processos em torno de um valor médio de um dado parâmetro ou de um conjunto de parâmetros (Pereira & Requeijo, 2012, p.347). Nesse seguimento surgem as técnicas para monitorizar o comportamento dos processos em relação às especificações dos seus parâmetros, de forma a controlar variabilidades, que podem ser inerentes ao processo ou causadas por agentes exteriores, e determinar se o processo tem ou não capacidade para produzir dentro das especificações pré-definidas.

Numa perspectiva mais tradicional do SPC, nomeadamente nas técnicas desenvolvidas por Walter Shewhart, na análise de parâmetros com características contínuas utilizam-se as cartas de controlo da média e da variabilidade de um processo, no caso de se tratar de amostras, ou caso contrário utilizam-se as cartas de observações individuais e da média móvel.

Uma carta de controlo tradicional consiste num gráfico onde o eixo das abcissas representa cada uma das amostras (ou das observações individuais), e o eixo das ordenadas representa os valores do parâmetro em estudo na carta. Nela existem três linhas horizontais paralelas: a linha central, a linha superior de controlo e a linha inferior de controlo. As duas últimas representam, respectivamente, o limite superior de controlo (LSC) e o limite inferior de controlo (LIC). A linha central corresponde ao valor esperado do processo, ou seja, ao valor médio do parâmetro em estudo (Usman & Kontagora, 2010). Um exemplo ilustrativo pode ser encontrado na Figura 3.4. Este exemplo contempla ainda linhas de aviso (superior e inferior), que são linhas que se encontram entre a linha central e a respectiva linha de limite, cuja finalidade é a de alertar que a amostra em estudo se encontra tão afastada do valor esperado que merece ainda assim uma chamada de atenção a quem estuda a carta.

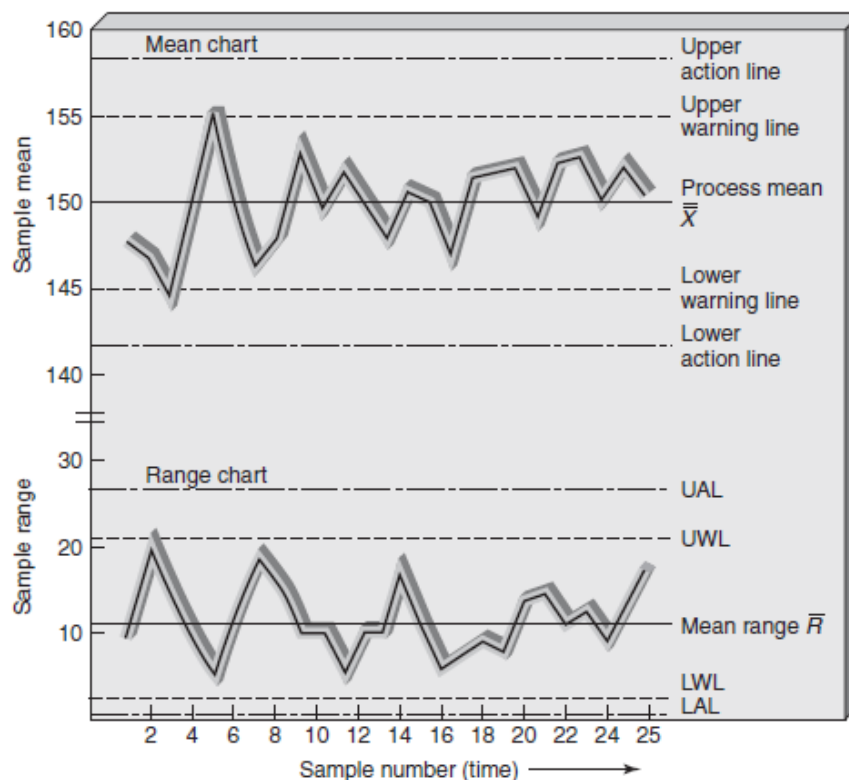


Figura 3.4- Exemplo de carta de média e amplitude. Fonte: Oakland, 2007

É possível ainda dividir uma carta de controlo em diversos “segmentos” ou zonas (Oakland, 2007, p.110), dentro dos limites de controlo. Considera-se um total de 3 zonas, a zona C (ou zona 1), que se conhece como zona estável e junto à linha central, a zona B (ou zona 2), a zona de aviso e a zona A (ou zona 3), a zona de acção, junto às linhas de limite de controlo inferior e superior. A Figura 3.5 é um exemplo disso mesmo.

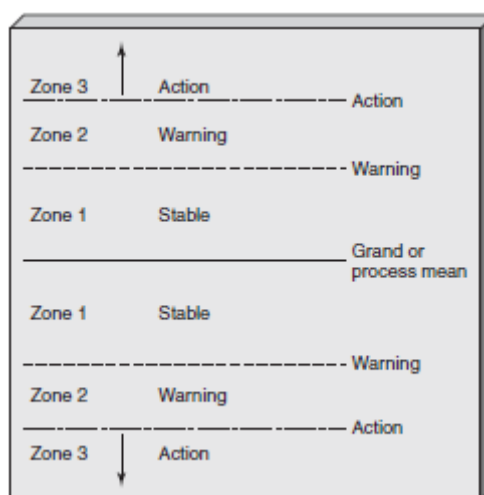


Figura 3.5 - As três zonas das cartas de controlo. Fonte: Oakland, 2007

3.5.1 Variabilidade e Causas de Variação

Nenhum processo tem a capacidade de produzir seja o que for sempre com os mesmos valores nos parâmetros. Para controlar essa variabilidade inerente aos processos surge então a necessidade de conhecer que tipos de causas a podem afectar. Existe então a distinção clara entre duas grandes causas de variação nos processos, que são as causas comuns de variação e as causas especiais de variação (Pereira & Requeijo, 2012).

Causas comuns de variação são causas aleatórias, criadas em consequência da variabilidade natural dos processos e dos seus inputs, sendo muito complexo determinar a sua causa, e sob as quais muito pouco pode ser feito para atenuar. Alguns exemplos são a vibração natural do equipamento, flutuações nas condições térmicas ou atmosféricas, variação de equipamento de teste e estado físico ou emocional dos operadores. A única forma de atenuação destas causas é com o envolvimento da gestão de topo, nomeadamente com equipamento de produção com capacidade de produção melhorada ou novos fornecedores (Oakland, 2007, p.69).

Por outro lado, as causas especiais de variação são características esporádicas e de magnitude superior às causas comuns de variação. Estas causas podem ser imediatamente identificadas nas cartas de controlo, devido à sua variação excessiva. Sempre que uma destas causas é detectada numa carta de controlo, a sua remoção imediata da carta de controlo é indispensável, normalmente por elementos mais próximos do processo (Pereira & Requeijo, 2012, p.350).

Uma definição extremamente importante no contexto das cartas de controlo é a de um processo estar sob controlo estatístico. Considera-se que um processo está sob controlo estatístico (ou “sob controlo”, ou ainda “estável”) quando este está sujeito somente a causas comuns de variação. Um processo onde existam causas especiais de variação considera-se um processo fora de controlo estatístico ou instável.

3.5.2 Detecção de causas especiais de variação

Na secção anterior refere-se que é simples detectar a existência de uma causa especial de variação numa carta de controlo, isto porque existem regras que permitem identificar determinados padrões seguidos pelos dados na carta que apontam para a existência destas causas.

A norma ISO 8258:1991 inclui oito regras, que são as mais usadas para a detecção de causas especiais de variação nas cartas tradicionais, também conhecidas como as cartas de Shewhart (Pereira & Requeijo, 2012, p.359). As regras são:

1. Qualquer ponto fora dos limites de controlo;
2. Nove pontos consecutivos na zona C ou para além da zona C do mesmo lado da linha central;
3. Seis pontos consecutivos no sentido ascendente ou descendente;
4. Catorze pontos consecutivos decrescendo e crescendo alternadamente;
5. Dois pontos em três possíveis consecutivos na zona A, ou do mesmo lado da linha central para além da mesma.
6. Quatro pontos em cinco possíveis consecutivos na zona B ou A, ou do mesmo lado da linha central para além das mesmas;
7. Quinze pontos consecutivos na zona C, acima ou abaixo da linha central;
8. Oito pontos consecutivos de ambos os lados da linha central, sem nenhum na zona C.

Se um ou mais destas situações se verificar, considera-se que o processo não está sob controlo estatístico, pois é detectada uma causa especial de variação. A regra mais básica para detecção destas causas é a regra 1, sendo que as restantes se destinam somente a aumentar a sensibilidade destas cartas face a possíveis padrões não aleatórios causados por causas especiais de variação.

3.5.3 Capacidade dos processos

Existe distinção entre um processo sob controlo estatístico e um processo capaz de produzir dentro de determinadas especificações para o produto/serviço. Caso se determine que um dado processo se encontra sob controlo estatístico, mas o mesmo é incapaz de produzir segundo especificações, é necessária intervenção da gestão de topo de modo a que essa situação se inverta, pois a capacidade do processo é função da variação causada pelas causas comuns (Pereira & Requeijo, 2012, p.409).

Uma qualquer especificação é geralmente definida pelo valor máximo e mínimo da tolerância, $+T$ e $-T$, respectivamente (no caso das especificações bilaterais). A relação entre a variabilidade do processo (sendo a mesma descrita pelo desvio padrão, ou σ) e a tolerância é a seguinte (Oakland, 2007, p.258):

- Alta precisão relativa: $2T \gg 6\sigma$;
- Média precisão relativa: $2T > 6\sigma$;
- Baixa precisão relativa: $2T < 6\sigma$.

Quer isto dizer que se considera que o processo será mais preciso quanto mais afastados estejam os 6σ de variabilidade do processo dos limites de especificação. Esta situação é ilustrada na Figura 3.6.

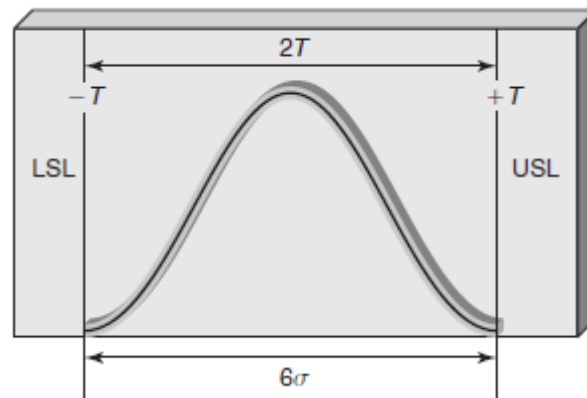


Figura 3.6 - Limites de especificação. Fonte: Oakland, 2007

De modo a determinar a capacidade de um processo de produzir segundo as suas especificações, definem-se índices de capacidade que determinam o potencial de capacidade do processo. São eles o índice de capacidade potencial (C_p) e o índice de capacidade (C_{pk}).

O índice de capacidade potencial compara a variação total do processo com as tolerâncias, e é dado pela equação (3.21).

$$C_p = \frac{LSE - LIE}{6\sigma} \quad (3.21)$$

onde:

LSE – Limite Superior de Especificação;

LIE – Limite Inferior de Especificação;

σ – Desvio padrão do processo.

É habitual considerar que o processo é potencialmente capaz se o valor de C_p for igual ou superior a 1,33 (Pereira & Requeijo, 2012, p.410).

Caso isso se verifique, não se deve concluir imediatamente que o processo é capaz, pois o mesmo pode ainda assim produzir material fora de especificação, pois este índice não avalia se o processo se encontra centrado relativamente à especificação técnica, ou seja, se a média do processo é coincidente com o valor nominal de especificação.

Daí surge o segundo índice, o chamado índice de capacidade, ou C_{pk} . O C_{pk} é o mínimo de dois outros índices, que são o C_{pki} e o C_{pks} . As suas expressões podem-se encontrar nas equações (3.22) e (3.23).

$$C_{pki} = \frac{\mu - LIE}{3\sigma} \quad (3.22)$$

$$C_{pks} = \frac{LSE - \mu}{3\sigma} \quad (3.23)$$

Considera-se que o processo é capaz quando C_{pk} , ou seja, o menor dos valores C_{pki} e C_{pks} , é superior a 1,33, e considera-se ainda que o processo tem a sua média centrada no valor nominal quando C_{pki} é igual a C_{pks} .

Caso se verifique que um processo está livre de causas especiais de variação, ou seja, sob controlo estatístico, e que o mesmo apresenta os índices de capacidade potencial e de capacidade superiores a 1,33, pode-se daí retirar conclusões sobre os parâmetros do mesmo e prosseguir com a aplicação do SPC, nomeadamente com a Fase II.

3.5.4 Fase I e Fase II do SPC

O processo de utilização das cartas de controlo serve dois objectivos distintos, mas complementares: Conhecer os parâmetros do processo em estudo e controlar o mesmo. Assim se divide a construção das cartas de controlo, ou seja, na Fase I tem-se como objectivo uma análise retrospectiva do processo, de modo a conhecer os parâmetros (média e desvio padrão, no caso das cartas de controlo de variáveis contínuas) do processo, que até aí eram desconhecidos, e na Fase II, cujo objectivo é, após conhecidos os parâmetros do processo, monitorizar o comportamento dos mesmos.

A Fase I termina quando o processo se encontra sob controlo estatístico, são conhecidos os parâmetros do processo e se determina que o processo tem capacidade. De seguida deve ser definido o plano de controlo e o tipo de carta a utilizar na Fase II do controlo, que decorre de forma contínua, monitorizando o comportamento do processo.

3.5.5 Medida de desempenho das cartas de controlo

Nas cartas tradicionais de Shewhart é assumido que os valores em estudo seguem uma distribuição normal, com média μ e variância σ^2 conhecidas. A performance da carta tradicional é muitas vezes medida em termos de uma variável discreta definida como *Average Run Length* (ARL) (Chakraborti, 2007).

O ARL de uma carta de controlo é definido como o número médio de amostras (ou observações) que devem constar na carta de modo a que se detecte um ponto fora de controlo, num processo sob controlo estatístico.

No caso das cartas Shewhart tradicionais, é comum definir-se o valor especificado para o ARL como sendo 370. Quer isto dizer que a um ponto que se encontre fora dos limites de controlo numa carta tradicional existe uma probabilidade associada do mesmo ser um “falso alarme”, sendo essa probabilidade de 0,27%. A essa probabilidade se chama a probabilidade de erro tipo I ou probabilidade de falso alarme, e é representada matematicamente pelo carácter α . Adoptando este valor para ARL, deve-se esperar que, em média, a cada 370 pontos exista um falso alarme, isto é, se detecte um ponto fora de controlo que não o é realmente (Pereira & Requeijo, 2012, p.354).

O objectivo de uma carta de controlo é o de ser mais precisa possível, ou seja, garantir um valor de ARL em controlo (ARL_{EC} ou $ARL_{Em\ Controlo}$) o maior possível e um ARL fora de controlo o menor possível (ARL_{FC}). O ARL_{FC} refere-se ao tempo esperado entre um processo ficar fora de controlo e esse evento ser detectado na carta de controlo. Maximizando o ARL em controlo diminuir-se-á o número de falsos alarmes, enquanto que minimizando o ARL fora de controlo é possível detectar mais rapidamente eventuais alterações aos parâmetros do processo.

3.5.6 Carta X e MR

Em situações onde a recolha de dados se cinge a observações individuais, as cartas de controlo tradicionais mais indicadas para controlo da dispersão (Pereira & Requeijo, 2012, p. 376) serão as cartas MR (sigla de *moving range*), uma variável fictícia criada, a partir do valor absoluto da diferença das observações consecutivas, de modo a estudar essa mesma componente, sendo que a média será controlada por uma carta X (ou seja, de observações individuais).

Apesar destas cartas serem menos sensíveis do que as restantes cartas de controlo tradicionais, dado o facto que controlam o valor da própria característica e não uma média ou desvio padrão de uma amostra, são uma boa solução para casos onde não seja possível a recolha de amostras para análise, como é o caso de, por exemplo, sistemas computadorizados de medição, processos químicos ou medições de características com técnicas destrutivas.

Os valores dos limites de controlo da carta X encontram-se nas expressões (3.24) a (3.26), e os da carta MR nas expressões (3.27) a (3.29).

$$LSC_X = \bar{Y} + \frac{3\overline{MR}}{d_2} \quad (3.24)$$

$$LIC_X = \bar{Y} - \frac{3\overline{MR}}{d_2} \quad (3.25)$$

$$LC_X = \bar{Y} \quad (3.26)$$

$$LSC_{MR} = D_4 \overline{MR} \quad (3.27)$$

$$LIC_{MR} = D_3 \overline{MR} \quad (3.28)$$

$$LC_{MR} = \overline{MR} \quad (3.29)$$

onde:

\overline{MR} – Média das amplitudes móveis;

\bar{Y} – Média dos valores observados.

Considerando que, no caso de a amplitude móvel ser obtida a partir de duas observações consecutivas, se tem $n=2$ e, portanto, $d_2 = 1,128$, $D_4 = 3,267$ e $D_3 = 0$. De notar que as expressões para os limites destas cartas aqui apresentadas só são válidas para situações onde os parâmetros do processo não sejam conhecidos *a priori*.

Nesta fase do estudo do processo, caso se detecte um ponto fora dos limites de controlo, o procedimento a realizar será retirar essa observação do estudo, sendo que os seus parâmetros deixam de ter contribuição para os parâmetros globais do processo, e nova observação deve tomar o lugar da observação retirada, mantendo constante o tamanho da amostra em estudo. Este procedimento só se deve realizar em situações onde os dados não sejam independentes, de forma a não alterar a dinâmica da correlação entre os mesmos.

Caso se considere que o processo está sob controlo estatístico, os parâmetros do processo da fase I, retirados a partir destas cartas, podem ser obtidos através das expressões (3.30) e (3.31).

$$\hat{\mu} = \bar{Y} \quad (3.30)$$

$$\hat{\sigma} = \frac{\overline{MR}}{d_2} \quad (3.31)$$

onde:

$\hat{\mu}$ - Estimador da média do processo;

$\hat{\sigma}$ – Estimador do desvio padrão do processo.

3.6 Controlo de pequenas produções

3.6.1 Short Runs

A interminável busca do mínimo custo e máxima eficiência exige que as organizações encontrem novas formas de aumentar a sua competitividade (Modarress *et al.* 2000). Com esse fim surgem novas técnicas de controlo da produção, nomeadamente a prática *Just-In-Time* (JIT), cujo ponto central foca-se na produção de pequenas quantidades. A isto se chama a produção em *Short Runs*, ou seja, a produção de um número reduzido de unidades num intervalo de tempo relativamente restrito, com nova produção a ocorrer findo um determinado período de tempo, ou em lotes específicos em número limitado.

Esta forma de gerir a produção cria novos desafios a nível do controlo estatístico da qualidade (Pereira & Requeijo, 2012, p.502), pois é impraticável a inspecção total do produto, existirá demasiada informação a analisar caso se criem muitas cartas de controlo para muitos produtos e a escassez de dados de cada produto dificultará a detecção de causas especiais de variação nas cartas e a rapidez de resposta às mesmas.

3.6.2 Cartas Q

Novas técnicas estatísticas aplicadas a este tema são desenvolvidas, com particularidades que permitem responder às dificuldades causadas por este novo paradigma de produção. Pereira & Requeijo (2012, p.502) referem as cartas de controlo de diferenças (Cartas D) e as cartas Z e W, para as situações onde existem dados suficientes para estimar os parâmetros do processo, e as cartas Q, CUSUMQ e EWMAQ no caso de não existirem dados suficientes para estimar os parâmetros do processo.

A aplicação das cartas Q neste estudo surge da necessidade de controlar um processo a funcionar com *short runs*, no qual existem dados insuficientes para estimar os parâmetros do processo (Pereira & Requeijo, 2012, p.529). São baseadas nas cartas de Shewhart tradicionais e mantêm o mesmo aspecto geral. A metodologia das cartas Q avalia as estatísticas de um dado instante tendo em conta os parâmetros estimados a partir de todas as observações até instante imediatamente anterior. É gerada uma estatística para cada observação, a que se chama a estatística Q de um determinado instante, que é função dos parâmetros estimados até ao momento.

Assumindo que os dados originais seguem uma distribuição normal (com dados independentes e idênticamente distribuídos), a estatística Q segue uma distribuição normal reduzida, com média nula e variância unitária. A expressão para a estatística Q encontra-se na equação (3.32).

$$Q_r(Y_r) = \Phi^{-1} \left(G_{r-2} \left[\left(\sqrt{\frac{r-1}{r}} \right) \left(\frac{Y_r - \bar{Y}_{r-1}}{S_{r-1}} \right) \right] \right), \quad r = 3, 4, \dots \quad (3.32)$$

$$S_r = \left(\frac{r-2}{r-1} \right) S_{r-1}^2 + \frac{1}{r} (Y_r - \bar{Y}_{r-1})^2, \quad r = 3, 4, \dots \quad (3.33)$$

onde:

Q_r – Valor da estatística Q no instante r;

Y_r – Valor da observação no instante r ;

Φ^{-1} – Inverso da Função distribuição normal;

G_v – Função de distribuição t-student com v graus de liberdade;

S_r – Desvio-padrão amostral no instante r , cuja expressão se encontra na equação (3.33);

Y_r – Observação no instante r ;

\bar{Y}_r – Média das observações até ao instante r .

Para controlar a dispersão do processo, tratando-se de observações individuais, utiliza-se a amplitude móvel como medida da variabilidade entre observações consecutivas, sendo que a partir da mesma se calcula uma estatística Q , cuja expressão se encontra na equação (3.34).

A mesma estatística só é calculada a partir do 4º instante, e posteriormente só em instantes pares para garantir que os dados são independentes e idênticamente distribuídos (segundo uma distribuição normal de média nula e variância unitária).

$$Q_r(MR_r) = \Phi^{-1} \left(F_{1,v} \left[\left(\frac{v(MR_r)^2}{(MR_2)^2 + (MR_4)^2 + \dots + (MR_{r-2})^2} \right) \right] \right), \quad r = 4, 6, \dots \quad (3.34)$$

com:

$$v = \frac{r}{2} - 1 \quad (3.35)$$

$$MR_r = |Y_r - Y_{r-1}| \quad (3.36)$$

onde:

v – Número de graus de liberdade, dado pela equação (3.35);

$F_{1,v}$ – Função de distribuição de Fisher, com 1 grau de liberdade no numerador e v graus de liberdade no denominador;

MR_r – Amplitude móvel no instante r , cuja expressão se encontra na equação (3.36);

Na eventualidade de se verificarem duas observações consecutivas com o mesmo valor, a expressão da estatística Q leva a uma indeterminação, causada pelo facto de $\Phi^{-1}(F_{1,v}(0)) = \Phi^{-1}(0) = -\infty$. Para ultrapassar esta questão define-se a estatística $W(MR)$, que embora não seja independente, assume todos os valores consecutivos a partir do 4º instante, se tal for necessário. A expressão da estatística $W(MR)$ é dada pela equação (3.37).

$$W_r(MR_r) = |Q_r(Y_r) - Q_{r-1}(Y_{r-1})|, \quad r = 4, 5, \dots \quad (3.37)$$

3.6.3 Limites de controlo para as cartas Q

Os limites de controlo das cartas Q, tanto para a média como para a dispersão são os seguintes:

- Limite superior de controlo (LSC) = 3;
- Limite inferior de controlo (LIC) = -3;
- Linha central (LC) = 0.

No caso das cartas W(MR), os limites são definidos por constantes que dependem do tamanho da amostra. Mas como se trata de amplitudes móveis de duas observações consecutivas, os limites assumem os seguintes valores:

- Limite superior de controlo (LSC) = $D_3 * d_2 = 3,685$;
- Limite inferior de controlo (LIC) = 0;
- Linha central (LC) = $d_2 = 1,128$.

3.6.4 Análise da capacidade do processo com as cartas Q

Embora se tenha garantido que o processo se encontra sob controlo e com capacidade de produzir segundo as especificações no estudo da fase I, ainda assim é necessário monitorizar na fase II a capacidade do processo, em tempo real, ou seja, à medida que os dados da produção se vão obtendo (Pereira & Requeijo, 2012, p.524).

Utilizando as cartas Q para controlar os processos produtivos exige que se controle a capacidade dos mesmos através de dois índices determinados a partir do tradicional C_{pk} , e são eles Q_I e Q_S . Estes dois índices são representados cronologicamente na carta Q, figurando em cada amostra o seu valor da estatística Q e os valores dos dois índices de capacidade nesse momento. As expressões para estimar os valores dos índices Q_I e Q_S para observações individuais podem-se encontrar nas equações (3.38) e (3.39), respectivamente.

$$(\hat{Q}_I)_r = \frac{LIE - \hat{\mu}_r}{k \hat{\sigma}_r} \quad (3.38)$$

$$(\hat{Q}_S)_r = \frac{LSE - \hat{\mu}_r}{k \hat{\sigma}_r} \quad (3.39)$$

sendo:

$$\hat{\mu}_r = \bar{X}_r \quad (3.40)$$

$$\hat{\sigma}_r = \frac{S_r}{c_4} \quad (3.41)$$

onde:

$\hat{\mu}_r$ – Estimador do valor esperado de um parâmetro no instante r;

$\hat{\sigma}_r$ – Estimador da variância de um dado parâmetro no instante r;

c_4 – Constante das cartas de controlo, cujo valor se pode retirar da tabela presente no anexo I.

k – Constante que depende do tipo de especificação do processo em estudo.

Os valores de k que é habitual considerar-se encontram-se na Tabela 3.2.

Tabela 3.2 - Valores de k em função do tipo de especificação

Tipo de especificação	Valor de k
Bilateral	1,33
Unilateral	1,25

Estes índices permitem ainda analisar se o processo se encontra a produzir de forma centrada ou não, através da simetria entre os índices Q_I e Q_S .

Caso num dado instante se detecte uma causa especial de variação, não fará sentido calcular o valor dos índices, nem incluir na estimação dos parâmetros futuros os valores do parâmetro actual que causou o ponto fora dos limites de controlo. Deve-se, isso sim, investigar a causa desse mesmo ponto fora dos limites e implementar medidas correctivas que se considere adequadas.

3.7 Cartas CUSUM e EWMA

3.7.1 Cartas de controlo especiais

Em complemento ou substituição das cartas tradicionais de Shewhart, são desenvolvidas cartas, chamadas cartas de controlo especiais, com uma característica muito particular: são cartas capazes de detectar pequenas alterações na média de um processo, e detectar de forma mais rápida essa mesma alteração, sem grande impacto nos custos (Pereira & Requeijo, 2012, p. 449). As cartas de Shewhart tomam em conta apenas as últimas informações que recebem sobre os processos, e como tal perdem sensibilidade a pequenas alterações no parâmetro do processo, situação que só é parcialmente corrigida com a alteração das regras ISO referidas na secção 3.5.2. A implementação das cartas CUSUM (Carta de Somas Acumuladas) e EWMA (Carta da Média Móvel Exponencialmente Amortecida) permite detectar essas pequenas alterações no parâmetro da média dos processos.

A carta CUSUM é construída através de duas semi-cartas, uma para detectar aumentos e outra para detectar diminuições da média do processo. Para esse fim são definidas variáveis que são função de determinados parâmetros que variam consoante a variação que se deseja detectar e o ARL que se espera da carta. Cada uma das semi-cartas pode ser representada no mesmo gráfico, e considera-se que o processo está fora de controlo estatístico quando a semi-carta que detecta o aumento da média é superior a um determinado parâmetro limite, e, segundo a mesma lógica, quando a semi-carta do decréscimo da média é menor que o simétrico desse mesmo parâmetro limite. Esta carta apresenta a vantagem de ser possível determinar o novo valor da média do processo, caso se detecte que o mesmo se alterou.

A carta EWMA é construída com as mesmas premissas da carta CUSUM, sendo que os seus limites de controlo são ainda função da variância da variável a ser controlada, além de um conjunto de parâmetros semelhantes aos da carta CUSUM. A estatística criada pela carta EWMA varia da carta CUSUM pelo facto da mesma ser exponencialmente amortecida, isto é, dá ponderações diferentes a cada observação anterior da mesma estatística.

Para controlo da dispersão dos processos, as cartas CUSUM e EWMA apresentam diversas variantes dependendo se se trata de amostras ou observações individuais, à semelhança das cartas de Shewhart tradicionais.

Tanto as cartas CUSUM como as EWMA apresentam ainda uma possível alteração que pode ser uma vantagem, em comparação com as cartas Q ou as cartas de Shewhart tradicionais, que é a possibilidade das mesmas terem uma atenção especial ao processo na sua fase inicial de controlo, ou seja, podem apresentar uma variação nos limites de controlo nos instantes iniciais de análise do processo, aumentando a sua sensibilidade nesses instantes. A estas técnicas chamam-se *Fast Initial Response* (ou FIR), e a sua construção é em tudo semelhante à carta análoga. Neste estudo apresenta-se uma introdução a cada uma destas duas cartas, mas opta-se por se desenvolver mais aprofundadamente a carta EWMA, pois a primeira não será aplicada no estudo prático.

A carta de controlo EWMA, na sua forma mais simples, é construída da seguinte forma: Assumindo que a variável em estudo segue uma distribuição normal, ou seja, com uma média μ_Y e variância σ_Y^2 conhecidas, define-se uma variável E que irá controlar pequenas alterações na média do processo. A expressão da variável E, para um instante t, é a descrita pela equação (3.42).

$$E_t = (1 - \lambda)E_{t-1} + \lambda Y_t, \quad t = 1, 2, \dots \quad (3.42)$$

onde:

E_t – Valor da estatística E no instante t;

λ – Constante de amortecimento;

Y_t – Valor observado da característica da qualidade em estudo no instante t.

Os limites da carta de controlo EWMA são dados em função da constante de amortecimento, do valor do $ARL_{Em \text{ Controlo}}$ desejado e da variância da característica em estudo. Os valores dos limites de controlo encontram-se nas expressões (3.43) a (3.45).

$$LSC_E = E_0 + K\sigma_E \quad (3.43)$$

$$LIC_E = E_0 - K\sigma_E \quad (3.44)$$

$$LC_E = E_0 \quad (3.45)$$

onde:

E_0 – Valor inicial da estatística E, igual à média do processo;

K – Valor do nível de controlo, para um determinado $ARL_{Em \text{ Controlo}}$ e λ ;

σ_E – Desvio padrão da estatística E, dada pela expressão (3.56).

$$\sigma_E^2 = \sigma_Y^2 \left(\frac{\lambda}{2 - \lambda} \right) (1 - (1 - \lambda)^{2t}) \quad (3.46)$$

A última parcela da equação anterior, dado que rapidamente tende à unidade, pode ser ignorada de forma a simplificar o cálculo da variância da estatística E, resultando em limites de controlo constantes, e não variáveis ao longo do tempo, situação que ocorreria se fosse considerada essa última parcela.

Para determinar os valores da constante de amortecimento e do nível de controlo, o primeiro passo necessário será definir um valor para o $ARL_{Em \text{ Controlo}}$ que se considere aceitável, e a alteração da média $\Delta = \delta \sigma_{\bar{X}}$ que se deseje detectar. A partir desses valores, e dos ábacos presentes no anexo II (ábacos EWMA) podem ser retirados os valores de K e de λ a considerar.

Nas cartas CUSUM e EWMA, considera-se que um processo está sob controlo estatístico caso não hajam pontos fora dos limites de controlo, não se aplicando as regras ISO referidas anteriormente, o que resulta em cartas mais simples de analisar que as cartas de controlo tradicionais.

3.7.2 Carta EWMA de resíduos

Como se referiu anteriormente, a utilização das cartas de controlo tradicionais, e consequentemente, das especiais (como a CUSUM e a EWMA) só é válida quando os dados são normalmente distribuídos e independentes.

Na presença de dados onde se verifique a existência de autocorrelação significativa, a utilização das técnicas anteriormente referidas só ocorrer para os resíduos ou erros de previsão das séries de dados, desde que os mesmos não apresentem correlação significativa, ou seja, possam ser considerados ruído branco (Pereira & Requeijo, 2012, p. 677).

A série dos resíduos de uma série temporal pode ser determinada de acordo com a expressão da equação (3.47).

$$e_t = Y_t - \hat{Y}_t \quad (3.47)$$

onde:

e_t – Resíduo da previsão no instante t;

Y_t – Valor da característica da qualidade no instante t;

\hat{Y}_t – Valor previsto da característica da qualidade no instante t;

Desta forma é possível aplicar a estatística E da carta de controlo EWMA à série dos resíduos, sendo válidas as expressões da secção 3.7.1, com a ressalva que o desvio padrão do processo deve ser substituído pelo desvio padrão dos resíduos ou do erro de previsão, a média do processo deve ser substituída por zero, que é o valor esperado dos resíduos ou dos erros de previsão) e que Y_t deve ser substituído por e_t .

3.7.3 Carta MCEWMA

A carta EWMA tem a sua utilização vedada a dados que não sigam uma distribuição normal, o que pode ser um problema em situações onde, quer os dados, quer os seus resíduos não apresentem essa característica. Para essa necessidade surge a carta MCEWMA (*moving centre*

line EWMA), que é capaz de controlar séries de dados não estacionárias ou nas quais se detecta auto-correlação.

Esta carta permite ainda não só o controlo do processo (a detectar causas especiais de variação) mas também a analisar a dinâmica do comportamento do mesmo (Pereira & Requeijo, 2012, pp. 685).

Apesar de ter uma construção relativamente complexa, a carta MCEWMA é de aplicação simples, pois o único input que exige, após a sua construção, é do(s) valor(es) observados da característica a controlar.

Dado que esta carta tem na sua génese a variável exponencialmente amortecida da carta EWMA, a expressão da estatística E é a mesma da carta anterior, e pode ser encontrada na equação (3.42).

Considerando que o valor previsto (\hat{Y}_t) da variável a controlar é igual ao ajustamento exponencial da estatística E realizado no instante anterior, o erro da previsão (e_t) no instante t será dado pela equação (3.48).

$$e_t = Y_t - \hat{Y}_t = Y_t - E_{t-1} \quad (3.48)$$

Os limites desta carta são dados pelas expressões (3.49) a (3.51).

$$LSC_t = E_{t-1} + 3\sigma_{ep} \quad (3.49)$$

$$LIC_t = E_{t-1} - 3\sigma_{ep} \quad (3.50)$$

$$LC_t = E_{t-1} \quad (3.51)$$

onde:

σ_{ep} – Desvio padrão dos erros de previsão.

De notar que os limites da carta vão-se alterando ao longo do tempo, em função da variável E, que por sua vez também sofre variações causadas pela evolução e pela auto-correlação da variável a controlar. É esta característica singular que permite analisar a evolução da característica em estudo na carta MCEWMA.

Como a carta MCEWMA é, na sua génese, uma carta de resíduos (apesar do seu input serem os dados brutos), esta carta apresenta menor sensibilidade para detectar pequenas variações da média do processo, o que não acontecia na carta EWMA. Para complementar esta carta de forma a reduzir essas limitações implementam-se índices auxiliares de controlo, conhecidos como *Tracking Signal*, definidos para cada instante de tempo. Um índice muito utilizado será o índice T_s , cuja expressão se pode encontrar na equação (3.52).

$$(T_s)_t = \left| \frac{A_t}{DAM_t} \right| \quad (3.52)$$

onde:

A_t – Erro amortecido no instante t , dado pela expressão (3.53);

DAM_t – Desvio absoluto médio no instante t , dado pela expressão (3.54).

$$A_t = (1 - \lambda_1)A_{t-1} + \lambda_1 e_t \quad (3.53)$$

$$DAM_t = (1 - \lambda_1)DAM_{t-1} + \lambda_1 |e_t| \quad (3.54)$$

onde:

λ_1 – Constante de amortecimento.

O valor do DAM para o instante inicial pode ser obtido, caso se conheça o desvio padrão dos erros de previsão (σ_{ep}), através da expressão da equação (3.55).

$$DAM_0 = \frac{\sigma_{ep}}{1.25} \quad (3.55)$$

No caso do erro amortecido, o seu valor inicial pode ser considerado nulo, e essa consideração deixa de ser significativa poucos instantes de tempo depois, devido à progressão geométrica gerada pela estatística A_t que depende do valor de λ_1 .

A nível de escolha dos parâmetros de amortecimento λ_1 (para as estatísticas DAM_t e A_t) e λ (para a estatística E), dois procedimentos distintos devem ser considerados. Para o primeiro parâmetro, λ_1 , o mesmo deve tomar um valor fixo não muito elevado, sendo aconselhado por Pereira & Requeijo (2012, p. 689) o valor de 0,03.

Para a constante de amortecimento da estatística E (λ) não se realiza o mesmo procedimento das cartas EWMA. Nesta situação deve haver lugar a um estudo dos erros gerados a cada instante para cada valor de λ , e escolher um valor para este parâmetro que minimize a soma quadrática dos erros e_t .

Nesta carta, considera-se que se detecta um ponto fora de controlo estatístico se uma das seguintes situações se verificar:

- Um ponto fora dos limites de controlo da carta MCEWMA;
- Três valores consecutivos da estatística T_s acima do valor crítico, que se considera igual a 0,5.

3.8 Utilização do SPC no controlo da qualidade da água

A estatística em geral é uma ferramenta da matemática muito presente no controlo da qualidade da água. Cun (1997) apresenta um estudo onde utiliza ferramentas estatísticas para caracterizar a água do rio Sena (Paris) ao longo de um período de 90 anos, onde analisa tendência e dispersão da série temporal dos dados de diversos parâmetros (como a presença de amoníaco, oxigénio dissolvido ou nitratos), aplicando com esse fim testes estatísticos como o teste de Bartlett, teste de Kendall, entre outros, e aplica ainda cartas cumulativas para representar concentrações de agentes em função do tempo de forma cumulativa.

Existem ainda trabalhos publicados (Cliff, 1983) que estudam a qualidade de análises de água, utilizando métodos estatísticos para determinar eventuais erros sistemáticos e detectar pequenas concentrações.

No campo do SPC existem diversos casos na literatura de utilização das cartas de controlo para o controlo da qualidade da água. Usman & Kontagora (2010) aplicam cartas de controlo tradicionais de Shewhart para analisar concentrações de químicos presentes em água engarrafada para consumo e respectivos impactos para saúde humana. Smeti (2007a) utiliza cartas de controlo tradicionais para controlar observações individuais em variáveis que não seguem distribuições normais. Existem ainda estudos de aplicação de cartas CUSUM no controlo de variáveis químicas em águas superficiais (Follador *et al.*, 2012) e de nascente (Zhou *et al.*, 2006), sendo que no último caso se aplica uma combinação de cartas Shewhart e CUSUM. Aizenchtadt (2008) propõe novas abordagens ao SPC, nomeadamente na fase II, no seu estudo de controlo de águas residuais, por considerar que o SPC tradicional não tem capacidade para responder no caso em estudo, introduzindo um elemento dinâmico ao estudo, onde os limites variam em função do tempo e da linearidade do sistema em estudo. Smeti (2007b) utiliza transformadas logarítmicas para ultrapassar dificuldades relacionadas com dependência e elevada variabilidade de dados para aplicar o SPC em determinados parâmetros químicos respeitantes à qualidade da água a distribuir por habitantes na Grécia.

O tema do controlo da qualidade é uma presença constante na “indústria” da água, sendo que utiliza todos os tipos de ferramentas que a estatística e a qualidade podem oferecer para garantir o cumprimento das especificações e o controlo dos processos de tratamento de água.

3.9 Algumas noções de dinâmica de fluidos

Nesta subsecção apresentam-se expressões que permitem determinar, por via teórica, o tempo que a água leva a percorrer uma conduta com troços de área de secção transversal e comprimento variáveis.

A volume de um fluido que passa por uma conduta, por unidade de tempo é conhecido como caudal volúmico, cuja unidade SI é o metro cúbico por segundo (m^3/s). A partir desta grandeza, é possível determinar o tempo que um fluido demora a percorrer um determinado segmento de tubo, dado pela expressão (3.56) (Paterson, 1983).

$$t = \frac{L}{Q/A} \quad (3.56)$$

onde:

t – Tempo que um determinado volume de água demora a atravessar um segmento de conduta;

L – Comprimento do segmento de conduta;

A – Área da secção transversal do segmento de conduta ocupada pela água;

Q – Caudal volúmico que atravessa o segmento de conduta.

A expressão anterior só é aplicável para Q constante. Caso contrário, deve-se alterar a expressão de modo a considerar o caudal volúmico como sendo uma função dependente do tempo. Essa situação encontra-se contemplada na expressão (3.57).

$$L = \frac{1}{A} \int_{t_{inicial}}^{t_{final}} Q(t) dt \quad (3.57)$$

Considerando $t_{inicial}$ como sendo igual a zero, o tempo que o fluido demorará a atravessar um segmento de conduta será igual a t_{final} .

Estendendo a expressão anterior para uma situação com múltiplas condutas de comprimento e/ou área diferente, e considerando que se encontram dispostos horizontalmente (isto é, sem efeitos gravíticos a considerar), pode-se chegar à expressão (3.58), mais geral para uma situação com j condutas em série.

$$L_1 A_1 + L_2 A_2 + \dots + L_j A_j = \int_{t_1}^{t_{final}} Q(t) dt + \int_{t_2}^{t_1} Q(t) dt + \dots + \int_{t_{inicial}}^{t_{j-1}} Q(t) dt \quad (3.58)$$

Considerando a função caudal como contínua e integrável no intervalo de tempo em estudo, a expressão anterior pode ser generalizada como se apresenta na expressão (3.59).

$$\sum_{i=1}^j L_i A_i = \int_{t_{inicial}}^{t_{final}} Q(t) dt \quad (3.59)$$

A expressão anterior permite que, conhecendo as dimensões dos troços e o comportamento da função caudal, a cada instante de tempo ($t_{inicial}$) se possa calcular o t_{final} , e saber o tempo que o fluido demorou a percorrer a conduta. Consequentemente, poder-se-á conhecer o caudal à saída da conduta nesse instante de tempo e as características físicas do mesmo (como, por exemplo, a condutividade).

Há, no entanto, uma ressalva a considerar: A expressão anterior deve ser utilizada com cuidado, pois não será tão correcta em situações com grande variação de caudal. No caso da água, como fluido incompressível, podem ocorrer situações onde, devido a um aumento do caudal, a água que entra num dado momento na conduta “empurra” o volume de água que nesse momento ocupa todo o comprimento da conduta, gerando uma onda de pressão que aumenta de forma quase imediata o caudal à saída. Considera-se ainda que a água ocupa toda a secção transversal da conduta, o que pode nem sempre se verificar.

Para fins do modelo, é considerado que, ao chegar ao reservatório, a água com diferentes características se mistura de forma imediata e homogénea, algo que na realidade não acontece. A forma como a água se mistura dentro de um reservatório depende da geometria do mesmo, de como é feita a entrada da água nesse mesmo reservatório, entre outros factores. Este pressuposto simplificará a análise do problema, mas será uma fonte de erro a ter em conta nos resultados obtidos pelo mesmo.

3.10 Qualidade da previsão

Quando se estuda o ajustamento de um modelo face a uma situação real, é necessário haver uma forma de avaliar a qualidade do mesmo, ou seja, estudar o erro do mesmo.

Em situações onde os dados seguem escalas semelhantes, um dos indicadores da qualidade do ajustamento que pode ser utilizado é o desvio médio absoluto (ou DMA) (Hyndman & Koehler, 2006). Este indicador, descrito na expressão (3.60), tem a mesma escala das unidades das séries em estudo, e como tal, pode ser um indicador fácil de compreender, além de simples de calcular,

tendo a desvantagem de ser mais sensível a *outliers*. Outro indicador que pode ser utilizado é o erro quadrático médio (ou EQM, descrito na expressão (3.61), embora a sua raiz seja mais popular, pois as suas unidades são as mesmas dos dados originais. Historicamente, ambos os indicadores têm sido muito populares na estatística, devido à sua fácil compreensão e aplicação.

$$DMA = \frac{1}{n} \sum_{i=1}^n |e_t| \quad (3.60)$$

$$EQM = \frac{1}{n} \sum_{i=1}^n (Y_t - \hat{Y}_t)^2 \quad (3.61)$$

onde:

n – Tamanho da série de dados;

e_t – Erro no instante t , ou diferença entre valor real e valor previsto nesse instante;

Y_t – Valor real no instante t ;

\hat{Y}_t – Valor previsto no instante t .

Outra forma de controlar, em tempo real, a qualidade da previsão será através da monitorização do termo do erro da modelação ARIMA. Sendo possível determinar a série dos desvios da previsão (isto é, a diferença em cada instante do valor previsto e do valor real) essa série poderá ser modelável, utilizando os modelos ARIMA mencionados anteriormente, de forma a ter um termo do erro do modelo, que será normalmente distribuído.

Esse termo do modelo pode ser colocado numa carta de controlo, de forma a determinar os parâmetros da média e da dispersão da sua série de dados, tornando possível o seu controlo através das cartas que se considerem mais adequadas.

3.11 Análise dos dados e estudo de *outliers*

Numa qualquer operação de recolha de dados, quer rotineira quer complexa, podem ocorrer observações que não correspondem exactamente à realidade. Estas observações anormais podem ser causadas por erros de julgamento ou de medição, do próprio método de medição, erros de computação de dados, ou outros motivos. Estas observações anormais são chamadas *Outliers* (Sachs, 1984, p. 279). A presença destes outliers pode ser indicação de variabilidade do processo, insuficiências do modelo ou dos dados, ou ambas as situações.

A presença de *outliers* num conjunto de dados pode influenciar as decisões que se tomam a partir desse mesmo conjunto, diminuindo a fiabilidade das conclusões. Dessa forma é crucial que essas observações sejam detectadas e medidas apropriadas para lidar com as mesmas sejam tomadas. A principal razão para detectar *outliers* é garantir a qualidade dos dados, e garantir que os mesmos representam com fidelidade a situação real. Mesmo *outliers* isolados podem ter um efeito na média e desvio padrão de uma grandeza, a ponto de causar com que a distribuição estatística gerada para explicar os dados já não represente a realidade (Kaya, 2010). No caso de dados univariadas, as variáveis são analisadas uma a uma, mas no caso de dados multivariados já deve haver uma atenção à interacção entre variáveis que pode influenciar as observações.

Kaya (2010) define que os *outliers* podem ser de dois tipos distintos:

- Tipo aditivo (*addictive outlier*, AO), ou *outlier* de tipo I;
- Tipo inovador (*Innovative outlier*, IO), ou *outlier* de tipo II.

O *outlier* do tipo aditivo (AO), é uma observação que ocorre em consequência de um qualquer erro de medição ou de registo, e sem consequências nas restantes observações.

O *outlier* do tipo inovador (IO) é uma observação que afecta as observações posteriores, ocorrendo devido à variabilidade natural dos processos. Ainda assim esta observação merece uma análise cuidada, pois pode ser indicadora de um processo de produção que necessita de um ajustamento.

A detecção de *outliers* tem inúmeras aplicações na indústria, como por exemplo na banca, prevenção de fraude electrónica, meteorologia, marketing, entre inúmeras outras.

Métodos clássicos de detecção de *outliers* são tanto mais poderosos quanto menor o número de *outliers* existentes num dado conjunto de dados, e é necessário ter em atenção que um *outlier* pode não ser apenas um “número errado”, mas sim uma parte integrante de um conjunto de dados que pode contribuir para explicar um dado comportamento. Sendo assim o analista deve ter cuidado a avaliar as causas que geraram o mesmo.

Existem diversos métodos estatísticos para detectar *outliers*, sendo alguns deles:

- Método gráfico: diagrama da caixa de bigodes ou método do interquartil;
- Teste de Dixon;
- Teste de Grubbs;

Os métodos anteriores, apesar de simples, apresentam a desvantagem da sua utilização ser restrita a conjuntos de dados independentes. Caso esse pressuposto não se verifique, outros métodos mais complexos devem ser utilizados. Box (1994) apresenta o método de detecção de *outliers* para dados correlacionados mais conhecido, pois iterativamente analisa todas as observações e detecta as que são anormais para os *outliers* AO e IO.

A análise da integridade de dados não se deve restringir apenas à detecção de *outliers*. Nesta dissertação utiliza-se também o *software* estatística para detectar falhas na recolha das séries de dados, como dados omissos. O *software* permite a estimação dos valores omissos, ou substituição de valores aberrantes por outros, que podem ser a média das observações adjacentes, ou calculados a partir da tendência da série. Neste campo deve reinar o bom-senso de quem analisa os dados.

Capítulo 4 – Metodologia

4.1 Introdução

Neste capítulo apresenta-se a metodologia seguida para a construção do modelo de previsão do valor da condutividade, começando pela análise e validação dos dados, passando para a construção do modelo, pela validação dos resultados obtidos pelo mesmo e a utilização do modelo para controlar o processo.

4.2 Análise dos dados de entrada

O sistema de telegestão implementado pelos SMAS de Almada permite controlar, on-line e em tempo real, o estado de qualquer rede do sistema. Quer isto dizer que é possível saber, minuto a minuto, os dados da condutividade e caudal em pontos estratégicos da rede (como reservatórios e captações subterrâneas) e o volume presente nos reservatórios. É deste sistema que são retiradas as séries temporais que vão servir de suporte à construção do modelo de previsão da condutividade.

O sistema de telegestão, estando ainda em fase de implementação, encontra-se com certas limitações, a nível de recolha de dados, que estão actualmente em fase de correcção. Estas limitações, a menos que corrigidas, vão ter impacto nos resultados obtidos através do modelo, pois o mesmo estará assente em dados que não correspondem à realidade. Entre estas limitações contam-se, a título de exemplo, falhas instantâneas nos dados, ou intervalos de tempo onde o sistema considera a condutividade numa dada conduta como sendo nula.

Utilizando o *software statistica*, estas falhas nos dados podem ser estimadas, sendo que o *software* constrói a série temporal original, e a partir da tendência da mesma estima a observação (ou observações) em falta. É crucial garantir que o modelo está assente em dados consistentes e sem falhas, pois qualquer inconsistência terá como resultado uma falha na previsão, uma previsão completamente desfasada da realidade ou mesmo um possível alarme de problema no processo dado pelo modelo quando nada de errado se passa na realidade (falso alarme).

O gráfico da Figura 4.1 mostra esse mesmo facto. O sistema tem diversas falhas, com maior repercussão em dois grandes intervalos de tempo no meio da série temporal, que são corrigidas pelo *software statistica*. Tratadas todas as séries de dados, o passo seguinte será a construção do modelo, tendo como base as séries de dados corrigidas.

Análise da série temporal de uma entrada numa rede de distribuição

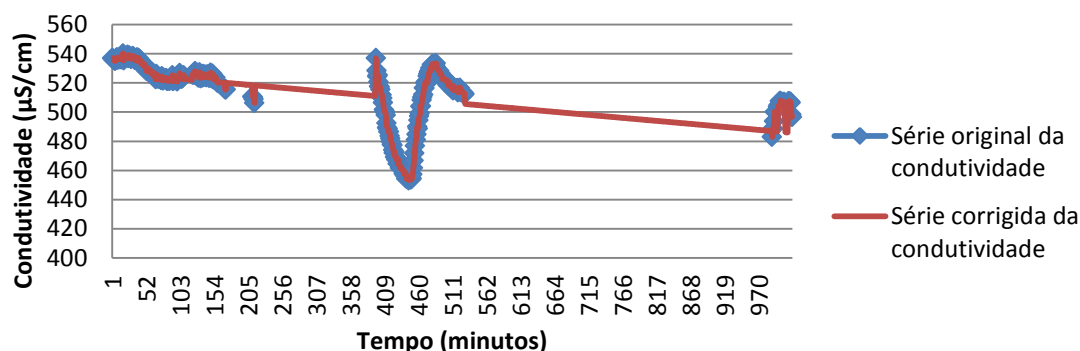


Figura 4.1 - Correção de uma série pelo *statistica*

Na figura anterior pode ver-se a azul os pontos existentes na série de dados original, e a vermelho as estimativas obtidas pelo *statistica* para os restantes pontos da série em falta, sendo que é esta última série que será utilizada nos modelos de previsão. Pode-se observar que os dados estimados seguirão a tendência da série real, mas são incapazes de reproduzir os picos que a mesma apresenta, e isto será uma fonte de erro a considerar no modelo a construir.

As restantes figuras representativas do tratamento das séries temporais utilizadas neste estudo podem-se encontrar no anexo II.

O caudal de cada uma das entradas do sistema também irá variar ao longo do tempo. Na prática, isto acontece porque há variações na procura de água ao longo do tempo, e, como tal, a necessidade de retirar água de um determinado furo subterrâneo nem sempre existe, e quando existe pode ter diferentes níveis de procura. O modelo terá que ter estas variações de caudal em conta. Na Figura 4.2 apresenta-se um gráfico da evolução do caudal volúmico que sai de um furo subterrâneo para o reservatório.

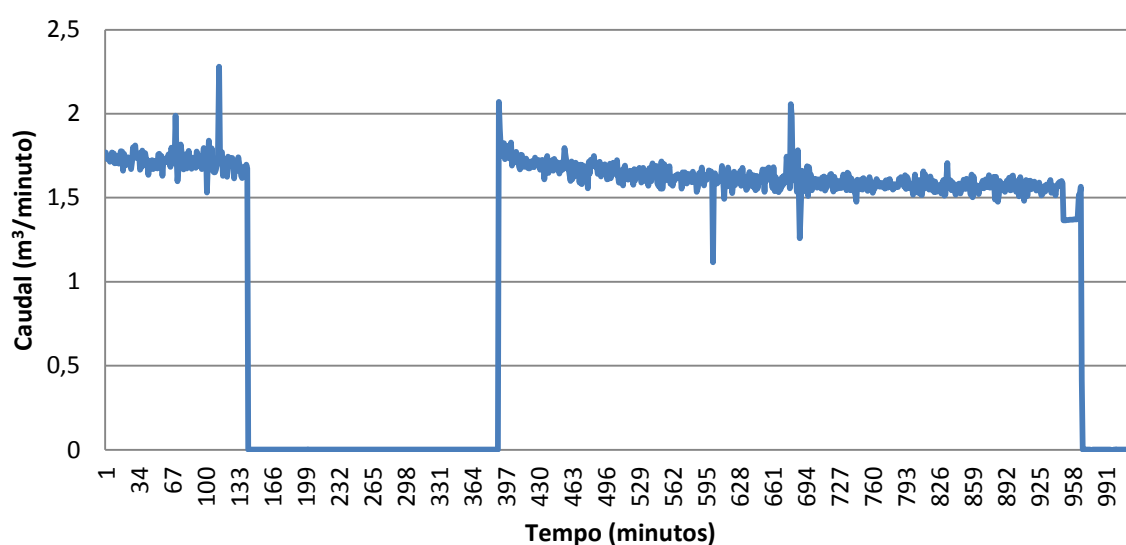


Figura 4.2 - Evolução do caudal de uma entrada do sistema

4.3 Construção do modelo de previsão temporal

Como foi referido anteriormente, o tempo deve ser uma variável a considerar no modelo. Ou seja, o mesmo deve ter em consideração o comprimento das condutas, a variabilidade da função caudal, e os seus efeitos no tempo que a água demora a atravessar um segmento de conduta.

Dessa forma, o algoritmo a construir deverá ser capaz de identificar, a cada instante, o tempo que a água demora a chegar ao reservatório vinda de cada fonte, como as condutas se interligam, e que efeitos isso pode ter na mistura de águas com características diferentes, tanto a nível de condutividade resultante como de caudal da mistura.

Já foi demonstrado, em estudos anteriores no SMAS de Almada, que a mistura de águas com condutividades diferentes segue a regra de uma média ponderada pelo volume (Silva, 2012), dado que as águas têm matrizes semelhantes. É com base nessa regra que o modelo irá fazer as misturas das águas provenientes de diversas fontes.

Um modelo simplificado pode-se encontrar, sob a forma de fluxograma, na figura 4.3. Este modelo contempla as misturas de águas provenientes de fontes distintas que se misturam apenas no reservatório final. No caso da existência de ligações entre as condutas, o modelo deve ser adaptado de forma a que se considere uma mistura das águas das duas fontes numa conduta, em vez de apenas uma fonte. Esta situação encontra-se no fluxograma da figura 4.4.

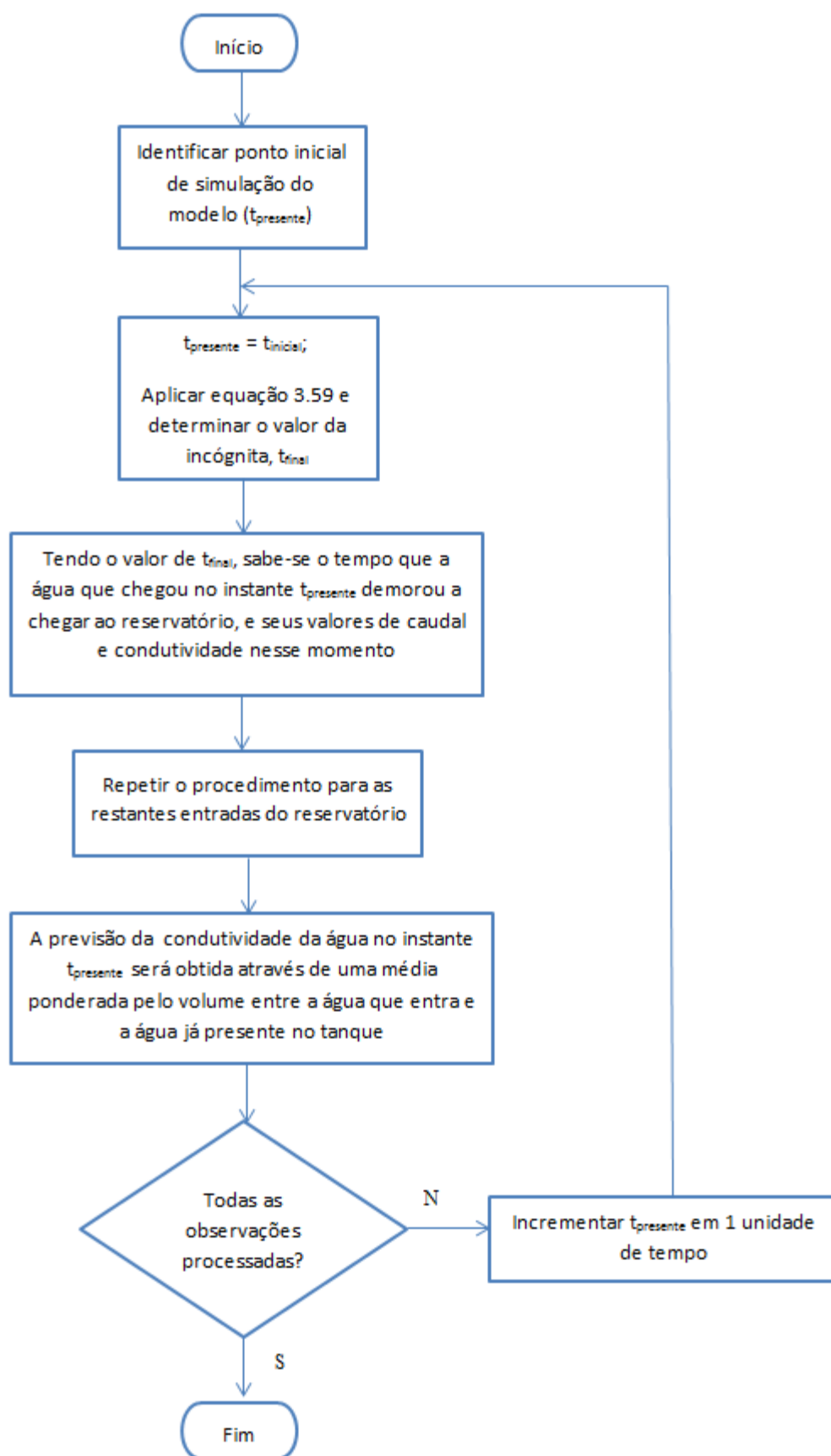


Figura 4.3 - Fluxograma do modelo simplificado

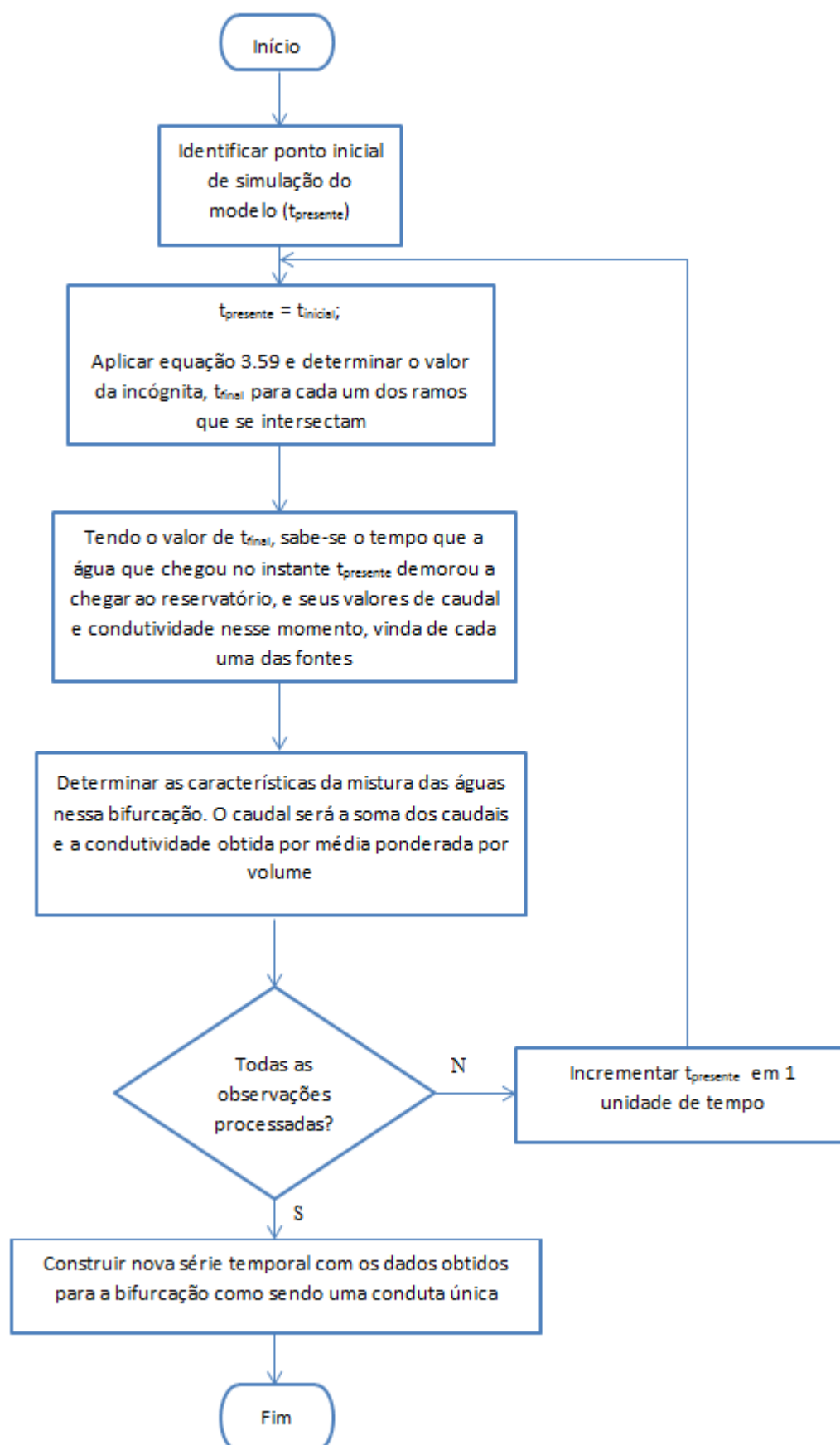


Figura 4.4 - Fluxograma de uma bifurcação

Combinando os dois fluxogramas é possível simplificar uma rede com bifurcações num sistema simples constituído por n inputs e um output, sendo que os inputs são as fontes de água que entram no reservatório, seu caudal e condutividade a cada instante de tempo, e o output será a previsão da condutividade da água no reservatório.

Para construir a ferramenta de *software* que irá fazer a previsão da condutividade no reservatório à saída do sistema foi utilizado o programa MATLAB. Neste *software* é possível construir um programa que seja capaz de analisar a função caudal de cada uma das entradas e ler as características da mesma, para todos os instantes de tempo, determinar as características da mistura das águas de diversas fontes e daí construir uma série temporal com os valores previstos da condutividade. Série essa que pode ser comparada com os valores reais que são observados dessa característica a cada instante, de modo a validar o modelo.

Para a construção dessa ferramenta no MATLAB, utiliza-se uma das redes existentes no concelho de Almada, cujo nome, localização e características específicas se encontram codificadas nesta dissertação. Um esquema simplificado da rede utilizada para construir o modelo encontra-se na Figura 4.5.

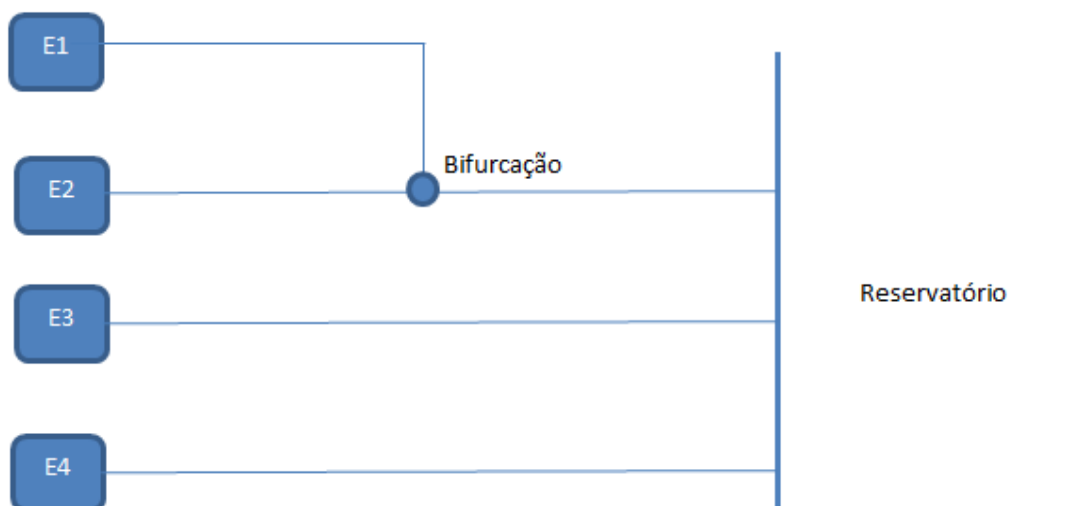


Figura 4.5 - Esquema simplificado de uma rede de abastecimento

Como se pode verificar na imagem anterior, a rede é constituída por 4 entradas de água, cada uma com características e caudais diferentes, e que variam ao longo do tempo. Existe ainda uma bifurcação entre a entrada 1 e a entrada 2. Esta situação terá que ser considerada no modelo, que fará a previsão da condutividade no reservatório aplicando os fluxogramas da figura 4.3 e da figura 4.4. Cada uma das ligações das entradas ao reservatório é constituída por diversos segmentos de tubagem distintos, onde varia o comprimento e área interna de cada secção.

Uma questão importante a referir é a natureza das entradas deste subsistema. As entradas identificadas por E2, E3 e E4 são captações subterrâneas, enquanto que a entrada E1 é outro reservatório de natureza semelhante ao reservatório de saída deste subsistema. Enquanto que ambos os reservatórios (de entrada e saída) dispõem de controlo de dados pela telegestão, permitindo conhecer as suas características em tempo real, o mesmo não ocorre nas captações subterrâneas, onde apenas é possível conhecer o seu caudal, não as características de qualidade

da água. As mesmas são controladas periodicamente pelos serviços dos SMAS de Almada. De forma a definir que valor da condutividade atribuir à água que provém de cada uma das captações subterrâneas recorre-se a um estudo anterior realizado no SMAS (Silva, 2012) que, utilizando cartas de controlo Q permitiu concluir um valor médio para essa característica, valores esses que se encontram presentes na Tabela 4.1.

Tabela 4.1 - Condutividades médias das captações subterrâneas do subsistema em estudo

Entrada	E1	E2	E3	E4
Condutividade ($\mu\text{S/cm}$)	Variável	1139.4	525.5238	508.875

A condutividade da entrada E1 que será inserida no modelo, a cada instante, será o valor real registado pela telegestão, enquanto que as outras entradas, sendo captações subterrâneas, terão um valor de condutividade que se considerará constante, e o valor inserido no modelo será o valor médio que se obteve no estudo em questão.

4.4 Modelo não temporal

Uma simplificação a fazer no modelo seria excluir totalmente a variável tempo da situação. Fisicamente, esta situação só faria algum sentido para extremas variações de caudal, algo que não acontece no sistema em estudo.

Neste modelo, retirar a variável tempo será o mesmo que afirmar que não existe tempo de passagem da água nas condutas, ou seja, que o instante em que a mesma é retirada dos reservatórios ou captações subterrâneas (de entrada) e em que é misturada no reservatório (de saída do sistema) é o mesmo. Para a previsão do valor da condutividade no reservatório de saída do sistema bastará então determinar o volume de água que sai num dado instante de cada captação e a sua condutividade, e determinar a condutividade da mistura das águas providas das diferentes entradas.

Este modelo será incapaz de reagir eficazmente a flutuações nos parâmetros de entrada, mas por outro lado é mais exacto quando as observações existentes são mais espaçadas no tempo.

Dado que o SMAS de Almada tem um sistema de telegestão que permite a obtenção de dados de minuto a minuto caso necessário, este modelo torna-se inadequado, daí a necessidade da variável tempo ser inserida no estudo.

4.5 Modelo híbrido

Para uma situação com variações de caudal relativamente pequenas, o modelo temporal apresentado anteriormente será uma boa representação do que se passa na realidade, mas numa situação com grandes variações de caudal em cada uma das entradas pode haver determinadas situações que este modelo não consegue contemplar.

Considerando, por exemplo, um brusco aumento de caudal num determinado instante. Num dado instante a água viaja através da rede a uma dada velocidade, e a que entra na rede no instante seguinte fá-lo com uma velocidade muito mais elevada. A água que flui na conduta, anteriormente com um valor inferior de caudal, será empurrada pela água que agora entra com

um caudal maior, dado que a água é um fluido incompressível e a área interna da conduta estar completamente preenchida pela água que flui. Este fenómeno propaga-se a uma velocidade grande pela conduta, e a água que se julgaria ter um determinado caudal, agora empurrada pelo caudal superior, terá um valor maior de caudal, o que trará erros a nível de cálculo da mistura das águas.

Para corrigir esse possível problema, propõe-se um modelo híbrido, ou seja, um modelo onde o tempo é factor para determinar as características da água no momento de entrada no sistema de condutas, mas o caudal a considerar será o do momento actual. Esta situação será apenas uma pequena alteração aos dois fluxogramas representados anteriormente.

Desta forma pretende-se minimizar os erros causados pela variação de caudal nas entradas que empurra a água que já estava nas condutas.

4.6 Controlo e monitorização da qualidade

Neste estudo utilizam-se cartas de controlo em dois momentos distintos. A primeira fase será para controlo da água recolhida nas captações subterrâneas, onde, devido ao reduzido número de observações (as recolhidas no estudo já realizado, com a aplicação das mesmas cartas Q) complementados com dados mais recentes, se utilizam novamente as cartas *short run* para o controlo da condutividade. Na segunda fase utilizam-se carta MCEWMA e EMWA para controlar a série da previsão da condutividade e para comparar o comportamento de cada uma das cartas.

Desta forma é possível ter um controlo da água que entra na rede, uma previsão das características que deverá ter à saída da mesma, e uma carta para controlar essa mesma previsão de saída. Desta forma é possível detectar eventuais falhas no sistema, quer nos inputs (causas especiais de variação nas cartas da entrada), quer na distribuição (através da previsão) e no reservatório final, com a previsão e as cartas que a controlam.

Capítulo 5 – Discussão de Resultados

Neste capítulo apresentam-se os resultados obtidos através dos diferentes modelos propostos para a previsão do valor da condutividade à saída da rede, os resultados obtidos a partir do controlo dessa mesma previsão, e o que se conclui acerca da qualidade do modelo desenvolvido.

5.1 Modelo temporal

O modelo temporal descrito no capítulo anterior foi aplicado, com recurso ao *software* MATLAB, para a rede específica descrita no capítulo anterior. Com os dados retirados da telegestão de minuto a minuto, através deste modelo, realizaram-se duas simulações distintas: uma com cerca de 500 observações reais, isto é, com os dados retirados do maior intervalo existente na série de dados com valores validados (que pode ser observado na Figura 4.1) e a outra com os dados estimados pelo *software estatística*, cuja estimação se encontra nessa mesma figura.

Os resultados obtidos, para a previsão do valor da condutividade da água, para cada uma das situações acima descritas encontram-se na Figura 5.1 e Figura 5.2.

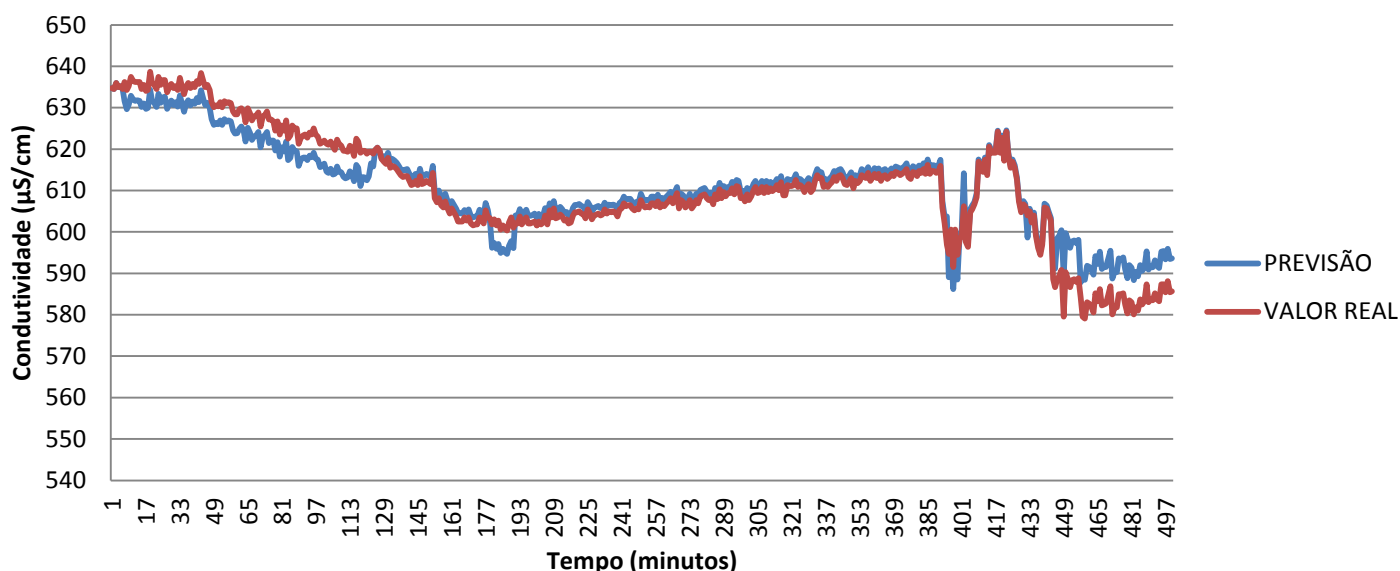


Figura 5.1 - Previsão da condutividade com dados reais

Na figura anterior, resultante da simulação com dados reais em 497 observações, pode-se observar que o modelo de previsão se ajusta bem ao comportamento do valor real da condutividade, existindo um período inicial onde o modelo ainda está em fase de ajustamentos, pois é impossível retirar dados anteriores ao início da série de dados, resultando nesse desfasamento. O modelo prevê os mesmos picos de condutividade no reservatório à saída, e tem um desfasamento no final da série causado não só pelas falhas nos dados que nesta situação ainda foram consideradas, como por factores externos ao modelo, tais como subestações secundárias a desviar volumes de água não contabilizados pela telegestão.

Na Tabela 5.1 resumem-se os valores obtidos para os parâmetros de avaliação do modelo de previsão, ou seja, o Desvio Médio Absoluto (DMA), o valor de desvio máximo do modelo face ao valor real, o Erro Quadrático Médio (EQM) e a sua raiz.

Tabela 5.1 - Parâmetros do modelo temporal com os dados reais

Parâmetro	DMA	Desvio máximo	EQM $(\mu\text{S}/\text{cm})^2$	\sqrt{EQM}
Valor $(\mu\text{S}/\text{cm})$	3.227	9.909	16.973	4.119

O passo seguinte será aplicar o mesmo modelo, mas desta vez considerando os inputs de dados como sendo os valores estimados através do *software statistica*, sendo que desta vez se simula com uma série de dados maior. Na Figura 5.2 pode-se encontrar o resultado dessa mesma simulação, com os valores da previsão e real da condutividade no reservatório de saída do sistema em estudo. Nesta série, tal como em todas onde se simula a partir dos dados estimados, a dimensão da mesma é de 1024 observações individuais.

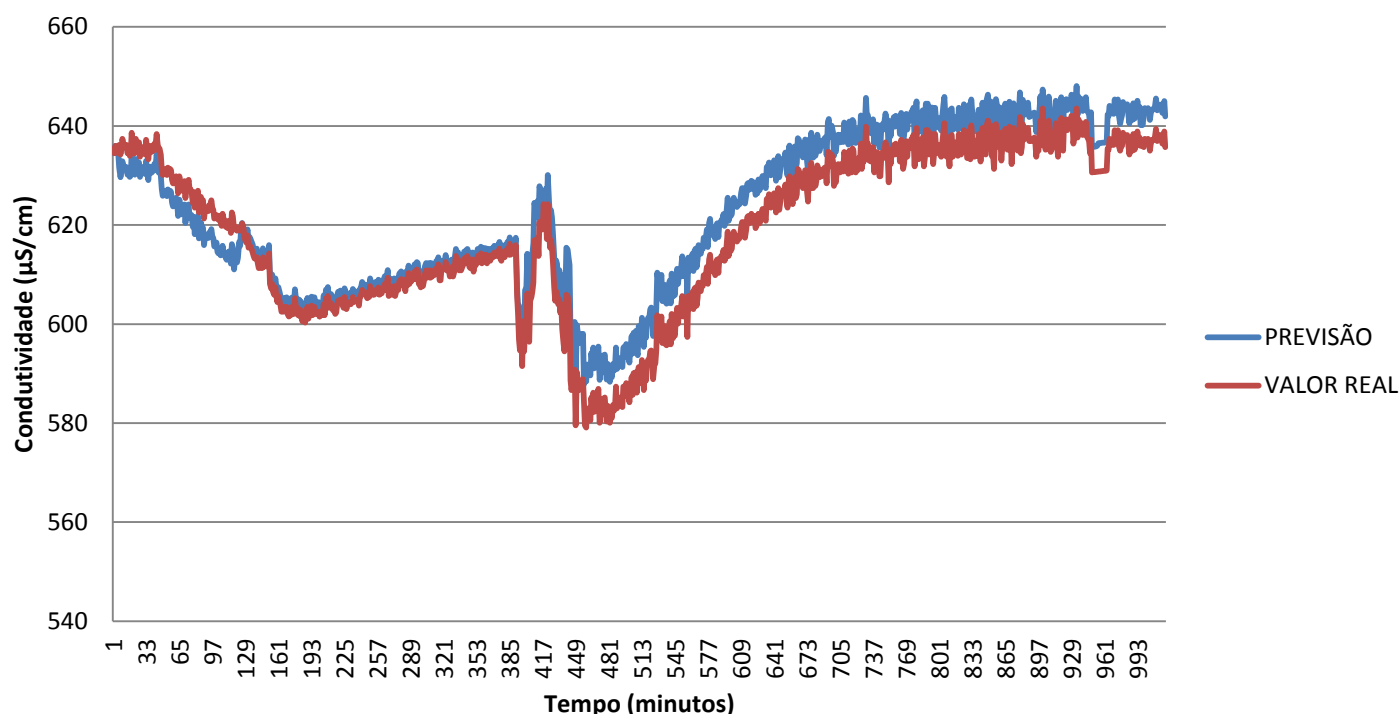


Figura 5.2 - Previsão e Valor real da condutividade num reservatório com inputs estimados

Esta simulação do modelo já foi executada com os inputs corrigidos, estimados a partir do *software statistica*. Pode-se observar a mesma falha inicial do modelo, e o ponto onde começa a adequar-se melhor à série de dados do valor real. Aproximadamente a meio da série de previsão, a mesma começa a sofrer de uma soma de padrão (continua a explicar a tendência da mesma forma, e representa em sintonia os picos da série real, mas com algumas unidades a mais em cada instante de tempo). Este desfasamento pode ser causado por factores como a estimativa dos inputs, desvio de água para subestações secundárias que não está contabilizado pela telegestão, mistura da água no reservatório de saída, entre outras.

Na Tabela 5.2 podem-se encontrar os valores dos parâmetros da previsão para este modelo.

Tabela 5.2 - Parâmetros do modelo temporal com os dados estimados

Parâmetro	DMA	Desvio Máximo	EQM ($\mu\text{S}/\text{cm}$) ²	$\sqrt{\text{EQM}}$
Valor ($\mu\text{S}/\text{cm}$)	4.883	9.909	29.555	5.436

As situações anteriores, como já foi referido, foram simuladas com recurso a um input de dados de minuto a minuto. Esta capacidade de recolha de dados é uma das maiores vantagens do sistema de telegestão implementado pelo SMAS de Almada, mas a capacidade do mesmo de retirar dados não se restringe por aí. Na realidade é possível retirar dados com qualquer periodicidade que se deseje, e, como tal, deve-se testar os modelos propostos para diferentes periodicidades de recolha de dados, de modo a poder daí retirar conclusões sobre qual o modelo se deve ajustar melhor a diferentes tipos de situações.

Nesse espírito, na Figura 5.3 apresenta-se o resultado da previsão obtido pelo modelo, para uma situação de recolha de dados a cada 33 minutos (aproximadamente). Estes resultados foram obtidos a partir de valores reais da condutividade, pois nesta modalidade de recolha de dados o sistema apresenta um número muito reduzido de falhas nas observações.

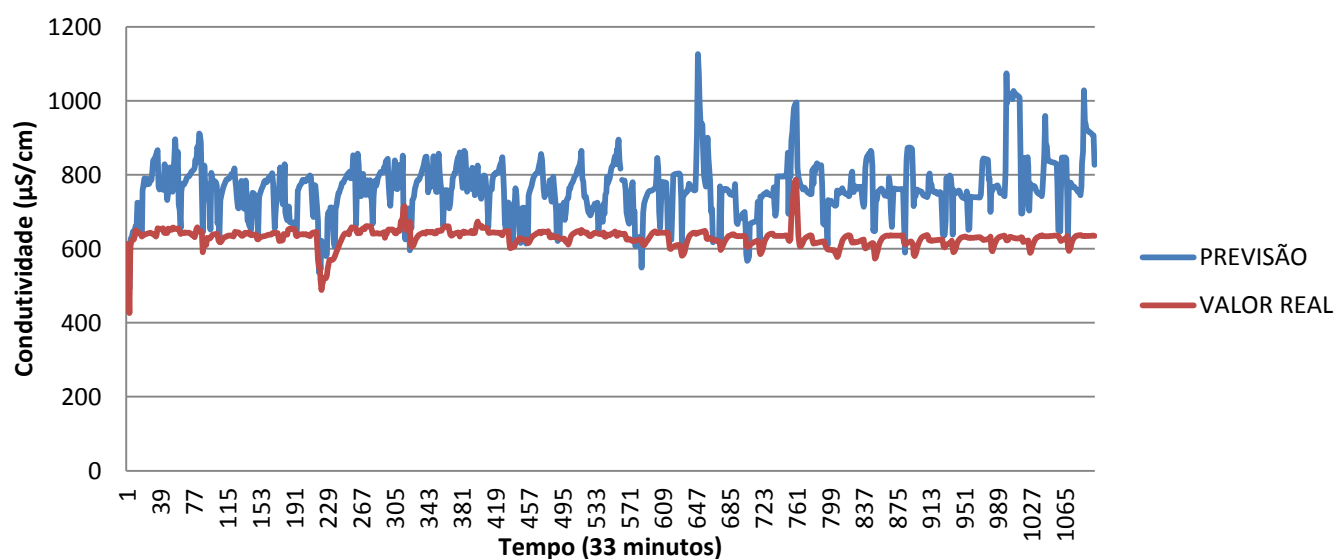


Figura 5.3 - Previsão e valor real da condutividade no reservatório com inputs reais a cada 33 minutos

Na figura anterior pode-se observar que, para esta periodicidade de recolha de dados a previsão é muito instável e claramente ineficaz, pois a série do valor real é relativamente estável, sendo a previsão incapaz de reproduzir esse comportamento. Na Tabela 5.3 resumem-se os parâmetros desta previsão, que como se pode ver pelo gráfico, serão inaceitáveis, pois a série do valor real apresenta um comportamento relativamente estável, ao invés da previsão, que se mostra demasiado sensível a pequenas alterações do valor real.

Tabela 5.3 - Parâmetros do modelo temporal com os dados reais com observações a cada 33 minutos

Parâmetro	DMA ($\mu\text{S}/\text{cm}$)	EQM ($\mu\text{S}/\text{cm}$) ²
Valor	131.8098	22158.88

Comparando com as duas simulações anteriores, o modelo para esta situação origina erros mais significativos, implicando que este modelo não será de todo aconselhável para situações onde o tempo entre as recolhas de dados seja mais elevado. Esta situação ocorre porque o tempo entre as recolhas de dados é demasiado elevado, pois em 33 minutos podem haver muitas variações de caudal ou de condutividade que o modelo é incapaz de incorporar, resultando nestes parâmetros de erros mais elevados do que nas restantes situações.

5.2 Modelo não temporal

Como já foi referido anteriormente, neste modelo a variável tempo é excluída, desconsiderando tempos de viagem de água através das diferentes condutas. Este modelo terá uma melhor capacidade de reproduzir os comportamentos da série do valor real da condutividade para situações onde a frequência de recolha de dados seja mais elevada, algo que o modelo anterior não foi capaz de responder adequadamente.

Na Figura 5.4 pode-se encontrar a evolução do valor previsto da condutividade no reservatório à saída do sistema e o seu valor real. Este modelo foi construído, à semelhança dos anteriores, com a série de dados completa estimada pelo *statistica*.

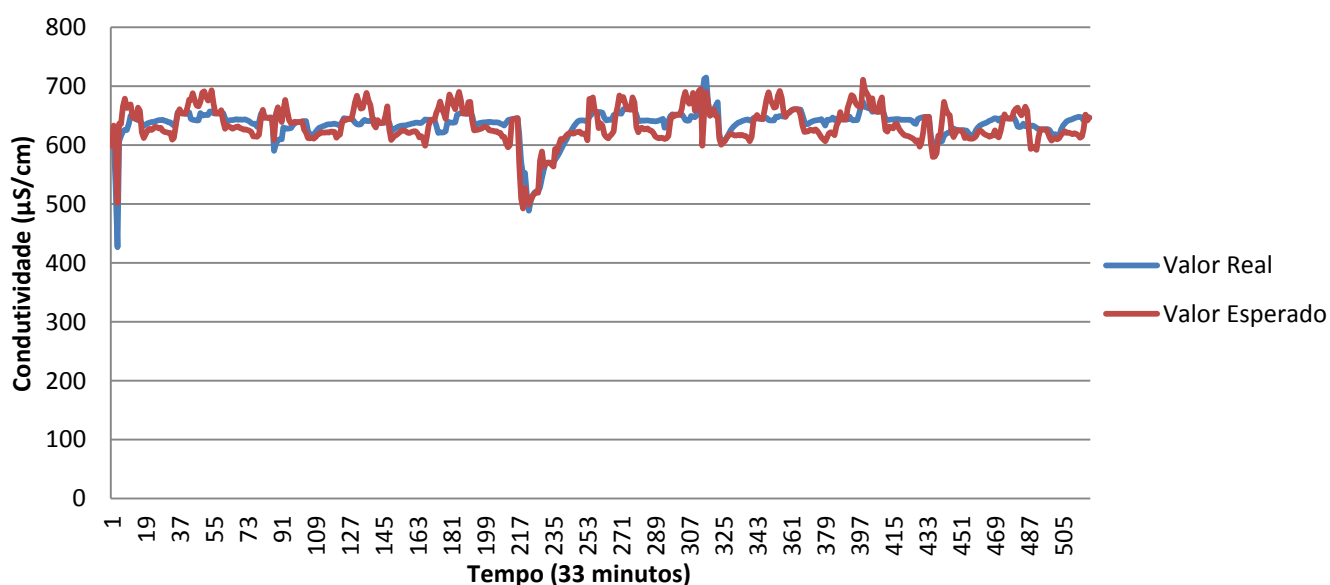


Figura 5.4 - Comparação entre valor previsto e valor real da condutividade no modelo não-temporal

Como se pode verificar, o modelo não-temporal já consegue acompanhar melhor as flutuações do valor real, sendo que ainda assim apresenta menos estabilidade que o mesmo, o que prova que este modelo não-temporal é mais facilmente influenciável por pequenas variações nos

inputs de dados. Na Tabela 5.4 encontram-se os resultados obtidos para os parâmetros do erro neste modelo.

Tabela 5.4 - Parâmetros do modelo não-temporal com observações a cada 33 minutos

Parâmetro	DMA	Desvio máximo	EQM $(\mu\text{S}/\text{cm})^2$	\sqrt{EQM}
Condutividade $(\mu\text{S}/\text{cm})$	17.128	88.279	488.179	22.094

5.3 Modelo híbrido

Neste modelo, como já foi referido anteriormente, misturam-se características únicas dos dois modelos anteriores, o temporal e não-temporal. Este modelo contempla a variável tempo na medida em que determina o tempo que a água demora a chegar ao reservatório de saída e daí tira a condutividade da mesma nesse instante, mas considera que a água foi entretanto empurrada pela água que posteriormente entrou no sistema, e considera o caudal da água no presente. Este modelo será mais correcto de aplicar para situações com grandes variabilidades nos caudais de entrada da rede.

Na Figura 5.5 pode-se encontrar o resultado obtido da aplicação do modelo híbrido, usando como input os dados reais não estimados pelo statistica, com observações de minuto a minuto.

Previsão e valor real da condutividade para o modelo híbrido com dados reais

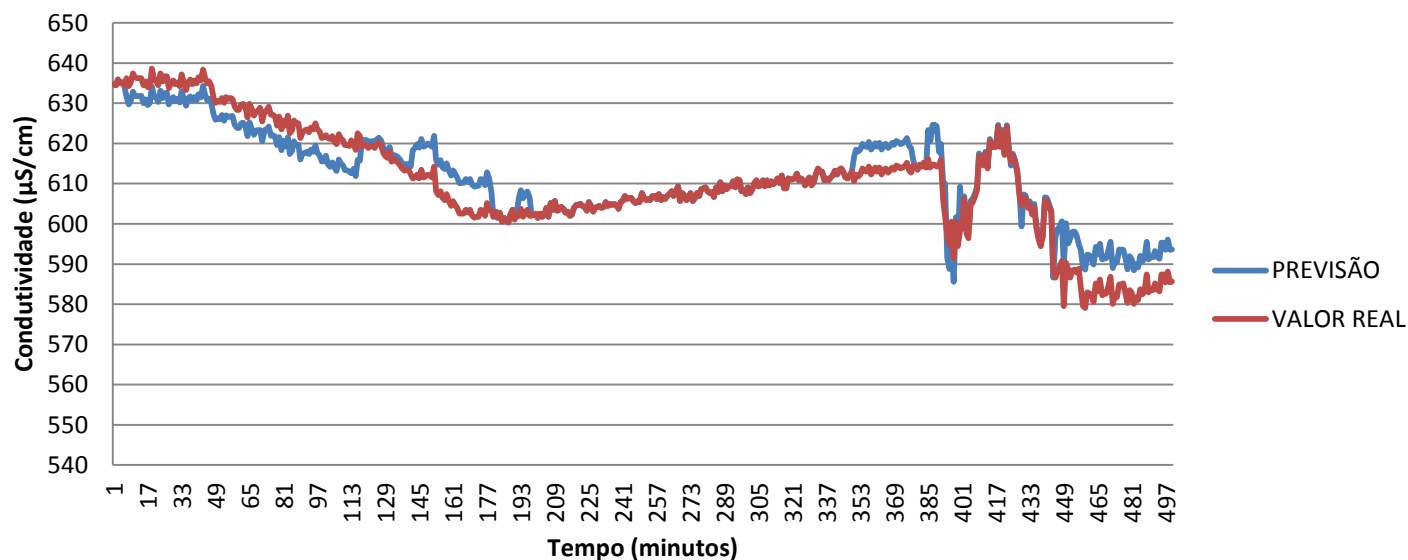


Figura 5.5 - Previsão e valor real da condutividade para o modelo híbrido com dados reais

Na Tabela 5.5 podem-se encontrar os parâmetros do erro obtidos com o modelo híbrido.

Tabela 5.5 - Parâmetros de erro do modelo híbrido com os dados reais

Parâmetro	DMA	Desvio Máximo	EQM $(\mu\text{S}/\text{cm})^2$	\sqrt{EQM}
Valor $(\mu\text{S}/\text{cm})$	3.390	10.836	22.576	4.751

Como nos casos anteriores, o próximo passo será aplicar novamente o modelo, mas agora com os dados extrapolados pelo *software statistica*. Na Figura 5.6 pode-se encontrar o resultado obtido da previsão da condutividade no reservatório de saída do sistema utilizando o modelo híbrido para os dados estimados.

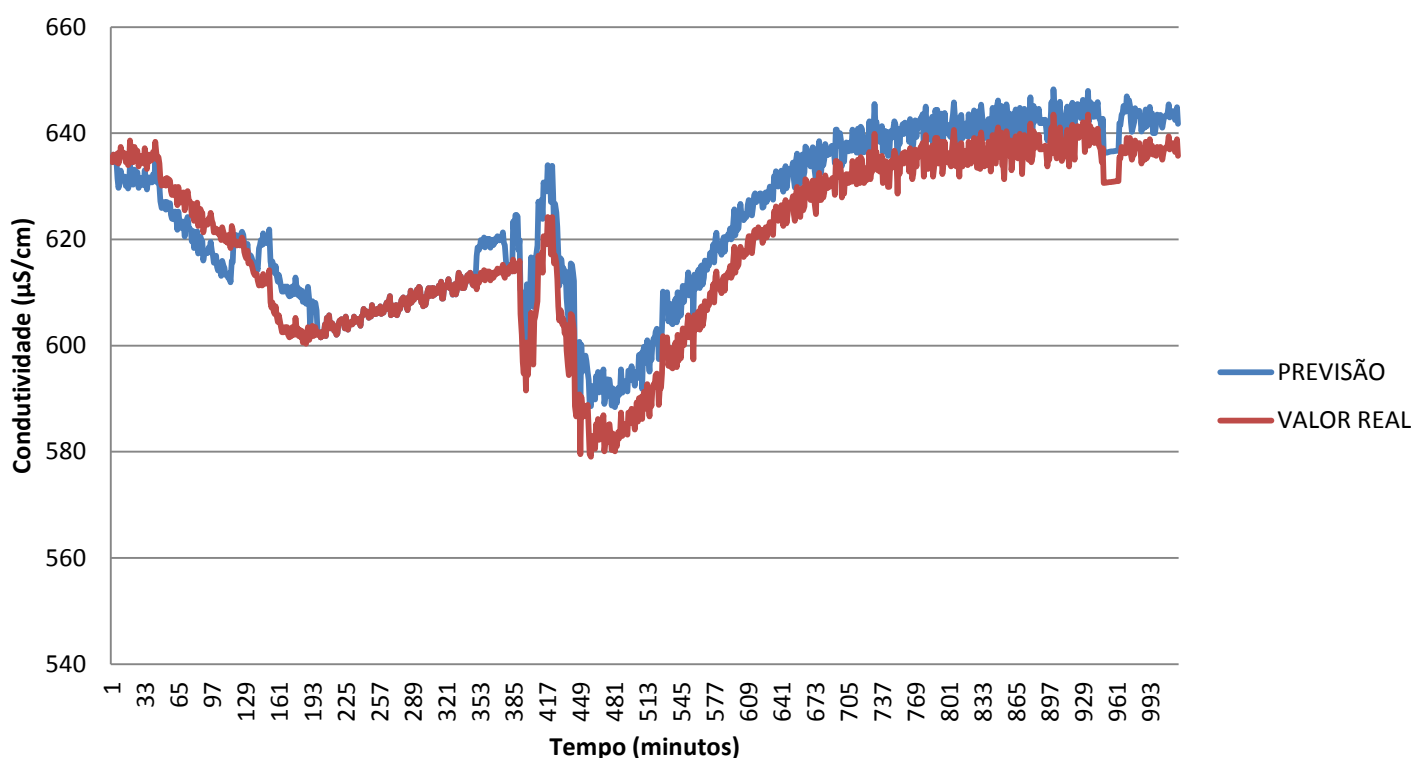


Figura 5.6 - Previsão e valor real da condutividade para o modelo híbrido com dados estimados

Este modelo devolve os parâmetros de erro que se encontram resumidos na Tabela 5.6.

Tabela 5.6 - Parâmetros de erro do modelo híbrido com os dados estimados

Parâmetro	DMA	Desvio Máximo	EQM $(\mu\text{S}/\text{cm})^2$	\sqrt{EQM}
Valor $(\mu\text{S}/\text{cm})$	5.238	11.247	32.869	5.733

5.4 Análise de sensibilidade do modelo previsional

Os modelos de previsão apresentados assentam muito fortemente em parâmetros de cada uma das entradas bem definidos, e eventuais variações em qualquer um deles que não esteja contemplada no modelo pode ter efeitos imprevisíveis na qualidade da previsão da condutividade.

Desta forma, para analisar a sensibilidade do modelo a flutuações dos parâmetros da entrada torna-se necessário alterar o valor da condutividade de cada uma das entradas de cada vez e observar o efeito que essa alteração terá (por exemplo) no desvio médio absoluto (DMA) da previsão.

Desta forma será possível observar qual das entradas terá maior influência nos valores observados da condutividade, nas quais se deve ter especial cuidado na monitorização dos parâmetros.

Na Tabela 5.7 apresentam-se os valores dos parâmetros obtidos para alterações da condutividade da entrada E2.

Tabela 5.7 - Sensibilidade do modelo a variações na entrada E2

Entrada E2 (todos os valores em $\mu\text{S}/\text{cm}$)					
Valor inicial da condutividade	DMA inicial	Alteração 1	DMA Alteração 1	Alteração 2	DMA Alteração 2
1139.4	3.23	1400	3.48	1600	3.79

Pode-se observar pela tabela anterior que quando o valor da condutividade se altera de 1139.4 para 1400 $\mu\text{S}/\text{cm}$, não resulta num aumento significativo no desvio médio absoluto da previsão, e o mesmo se o aumento for para os 1600 $\mu\text{S}/\text{cm}$.

Realizando o mesmo procedimento para a entrada E3, obteve-se os resultados sumarizados na Tabela 5.8.

Tabela 5.8 - Sensibilidade do modelo a variações na entrada E3

Entrada E3 (todos os valores em $\mu\text{S}/\text{cm}$)					
Valor inicial da condutividade	DMA inicial	Alteração 1	DMA Alteração 1	Alteração 2	DMA Alteração 2
525.52	3.23	700	4.16	1000	5.79

Nesta entrada pode-se verificar um aumento mais significativo do desvio médio absoluto da previsão, acompanhando a subida de 525 para 1000 $\mu\text{S}/\text{cm}$. Nesta entrada verifica-se que quando a condutividade for o dobro da esperada pelo modelo, o erro também será aproximadamente o dobro.

No anexo II encontram-se os resultados obtidos pelo mesmo estudo para a entrada E4. Na mesma não se verificam grandes alterações no desvio médio absoluto da previsão à medida que aumenta a condutividade da entrada.

O caso da entrada E1 é diferente. Ao passo que as entradas E2, E3 e E4 são captações subterrâneas, sem dados da telegestão, e como tal, o modelo recebe delas apenas o valor constante da condutividade obtido através das medições trimestrais e das cartas que as controlam, a entrada E1 é um reservatório semelhante ao reservatório de saída da rede que se

estuda nesta dissertação, e como tal, existem dados da telegestão para os parâmetros dessa mesma entrada.

Desta forma não será possível um aumento fixo para um determinado valor do parâmetro para o estudo de sensibilidade, mas é possível somar ao valor da condutividade um determinado valor a cada instante, de modo a poder simular eventuais fontes de poluição que corrompem a qualidade da água, e estudar como isso afecta a qualidade da previsão.

Na Tabela 5.9 - Sensibilidade do modelo a variações na entrada E1 encontram-se os resultados obtidos para a qualidade da previsão através da soma de determinados valores à condutividade da entrada.

Tabela 5.9 - Sensibilidade do modelo a variações na entrada E1

Entrada E1 (todos os valores em $\mu\text{S}/\text{cm}$)					
Valor inicial da condutividade	DMA inicial	Alteração 1	DMA Alteração 1	Alteração 2	DMA Alteração 2
Variável	3.23	300	6.14	650	9.81

Como se pode verificar, aumentando a condutividade de cada instante em 300 $\mu\text{S}/\text{cm}$ o desvio médio absoluto da previsão aumenta em aproximadamente 100%, e caso haja uma alteração nessa entrada de 650 $\mu\text{S}/\text{cm}$ em cada instante de tempo em relação àquilo que o modelo previa, o desvio médio absoluto da previsão aumenta em cerca de 200%. Pode-se observar então que os parâmetros da previsão são muito sensíveis a eventuais variações desta entrada em particular, e como tal a mesma deve ser alvo de um controlo mais apertado.

5.5 Cartas de controlo *short run* nas captações

Para controlar a água que entra no sistema através das captações subterrâneas aplicam-se as cartas de *short run*, mais especificamente a carta Q, porque o número de observações nas captações é relativamente reduzido, e não é constante de captação para captação.

Na Figura 5.7 apresenta-se o gráfico da série de dados recolhida da condutividade da entrada E4.

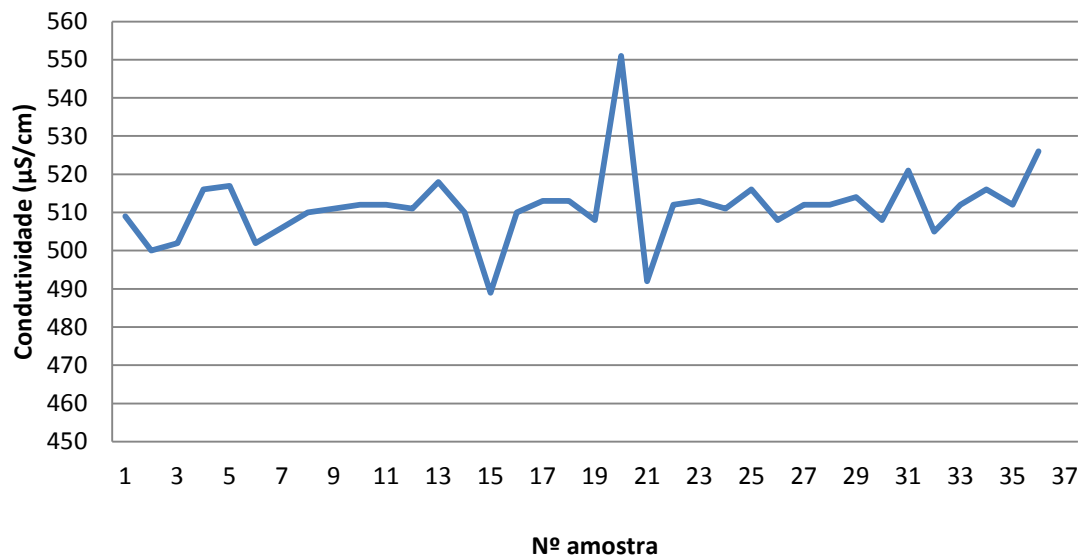


Figura 5.7 - Condutividade da entrada E4

De seguida apresentam-se as cartas Q da entrada E4 do sistema em estudo. Na Figura 5.8 pode-se ver a carta para o controlo da média da condutividade.

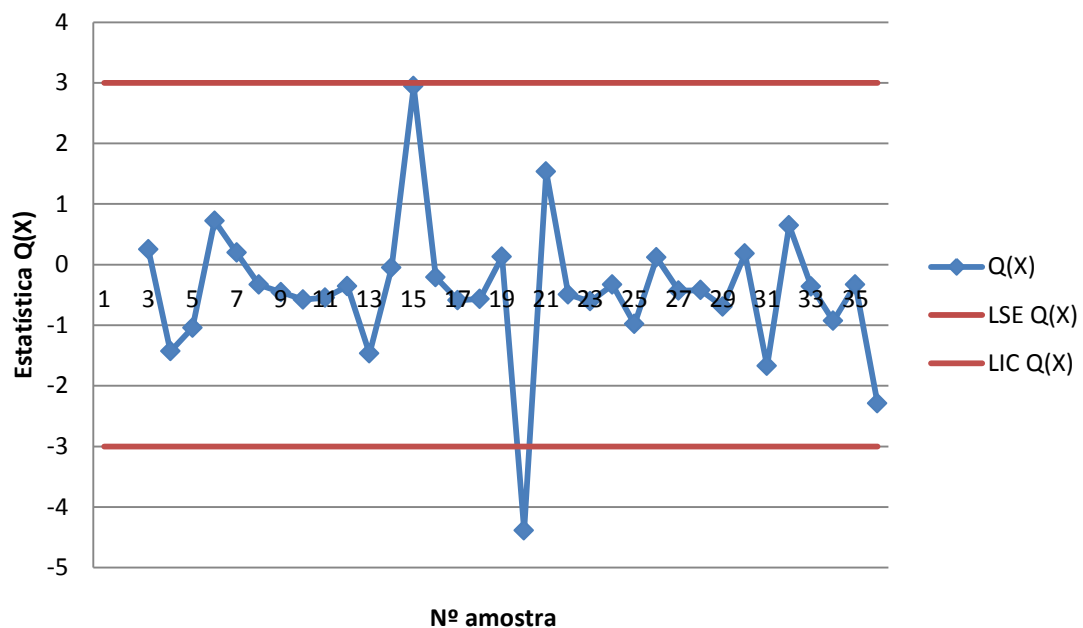


Figura 5.8 - Carta Q(X) da entrada E4

O passo seguinte será o controlo da dispersão, através da estatística Q(MR), cuja carta se apresenta na Figura 5.9

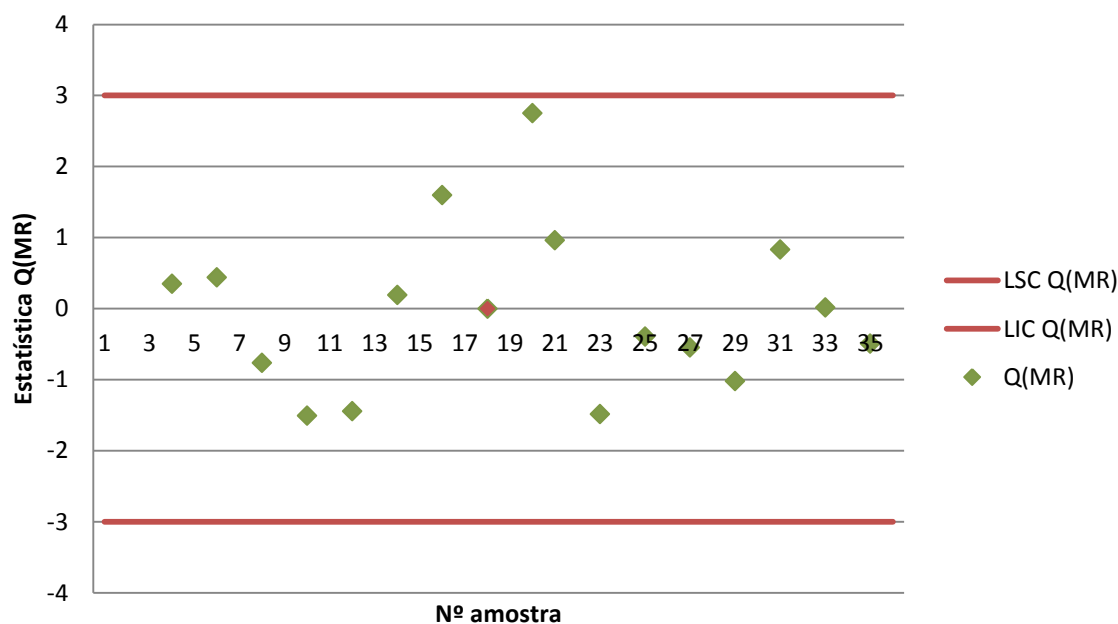


Figura 5.9 - Carta Q(MR) da entrada E4

Na carta que controla a dispersão ocorre uma indeterminação, pois verificaram-se duas observações consecutivas com o mesmo valor (observação nº18 da carta Q(MR), representada a vermelho e com o valor nulo atribuído automaticamente). Dada esta situação recorreu-se à carta W(MR) para complementar o controlo da variabilidade da condutividade da água nesta entrada. A carta W(MR) pode-se encontrar na Figura 5.10.

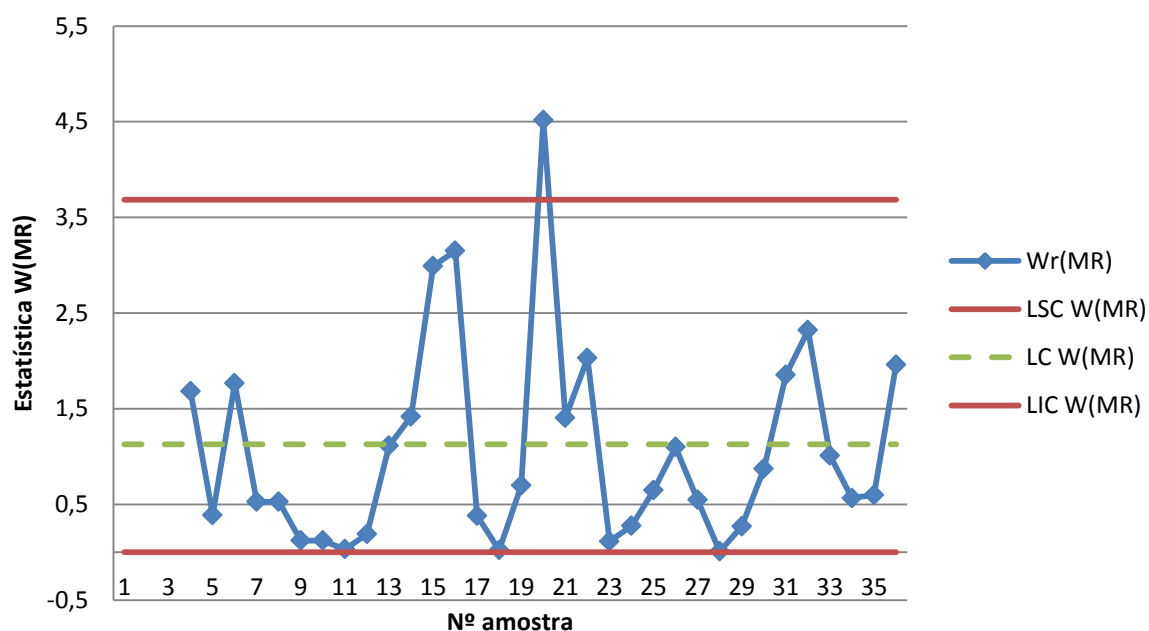


Figura 5.10 - Carta W(MR) para a entrada E4

No caso da entrada E1 verificou-se uma situação que merece um estudo mais aprofundado. O gráfico da sua série de dados da condutividade apresenta-se na Figura 5.11.

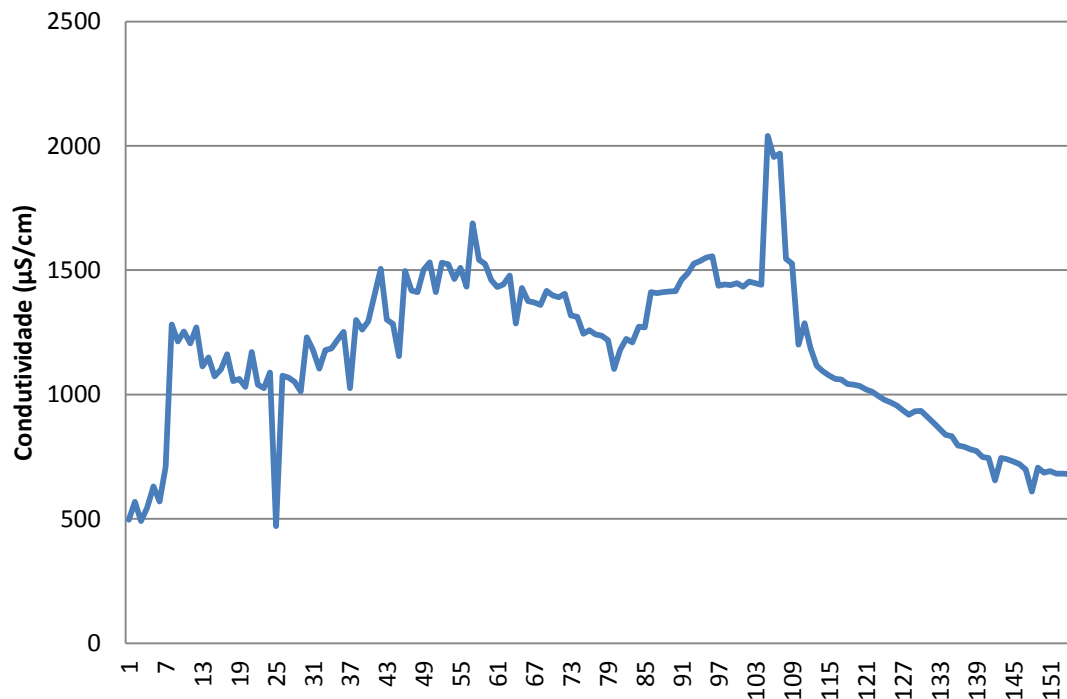


Figura 5.11 - Série de dados da condutividade da entrada E1

Ao contrário da entrada E4, que apresenta uma série muito mais suave, esta entrada E1 tem uma variação muito mais acentuada, e apresenta um intervalo de valores bastante maior do que a entrada E1. Esta situação implica que esta entrada funciona por “patamares”, isto é, toma um determinado intervalo de valores mais ou menos restrito num intervalo de tempo, e noutro intervalo de tempo apresenta outra gama de valores completamente distinta. Isto é resultado de um processo natural, e não é, de todo, implicativo que o processo esteja a “funcionar” de forma incorrecta. De forma a corrigir esta situação aplicam-se as cartas de controlo ao último “patamar” de valores da entrada, restringindo o tamanho da amostra ao patamar de valores mais recentes. Desta forma pode-se aplicar as cartas Q a um conjunto mais restrito de dados da condutividade desta entrada. A carta Q de controlo da média apresenta-se na Figura 5.12. Esta carta foi construída com base nas últimas 34 observações do gráfico anterior.

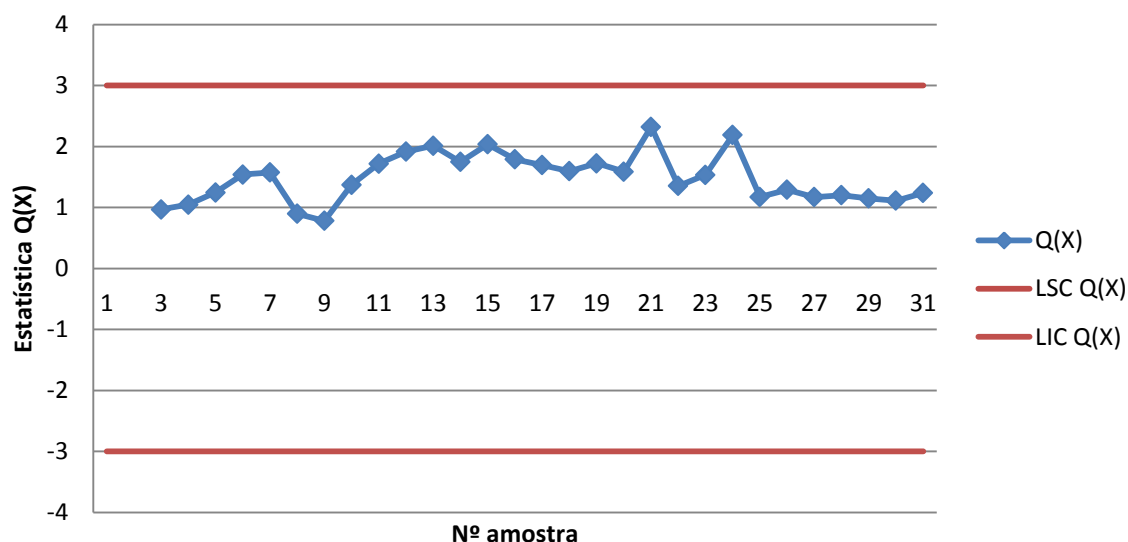


Figura 5.12 - Carta Q(X) do último patamar da entrada E1

Na Figura 5.13 apresenta-se a carta de controlo da variabilidade (Carta Q(MR)) do último patamar de valores da condutividade da entrada E1.

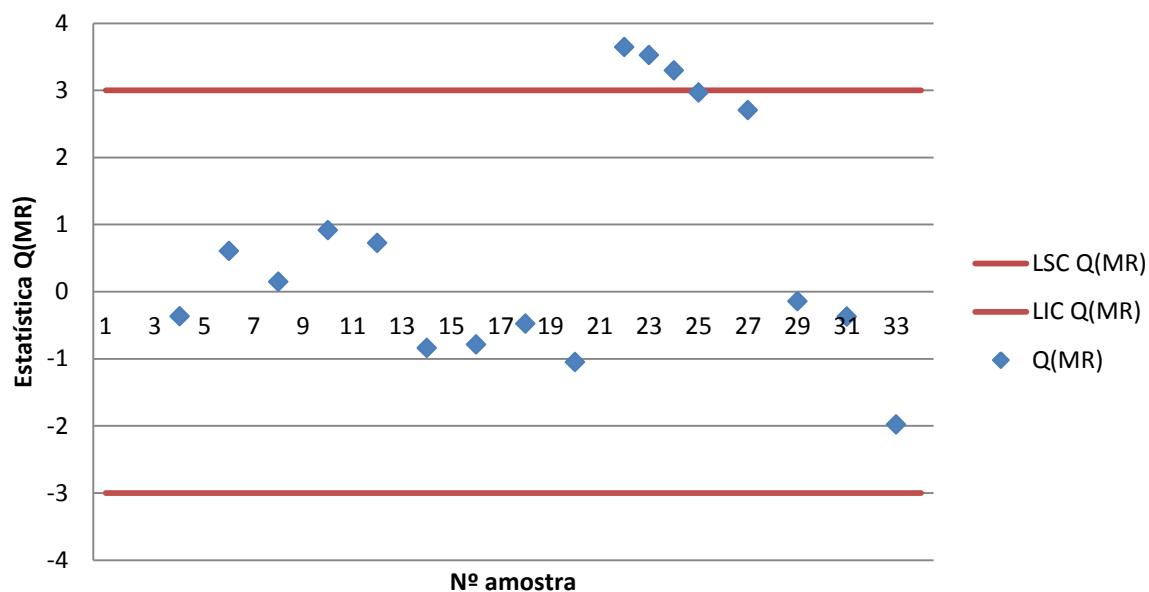


Figura 5.13 - Carta Q(MR) do último patamar de valores da condutividade da entrada E1

Apesar de se considerarem 34 observações para a construção das cartas anteriores, apenas 31 se encontram nas cartas, dado que 3 delas se encontravam fora dos limites de controlo da carta de controlo da variabilidade. Este facto vem atestar a extrema variabilidade que se encontra nos valores da condutividade da entrada E1, muito mais sensível a flutuações do parâmetro do que as restantes entradas.

As cartas de controlo da entrada E3 encontram-se no anexo III.

5.6 Controlo em tempo real da qualidade da previsão do modelo

Os parâmetros de qualidade do ajustamento do modelo da previsão, face ao valor real, embora úteis no que toca a comparar diferentes modelos, não permitem um controlo em tempo real da qualidade da previsão, e não permitem detectar num determinado instante a existência de um desvio da previsão que indicaria um problema de qualidade na rede. Como tal, torna-se necessário definir uma forma de controlar a previsão do valor da condutividade, face ao valor real, de forma a detectar eventuais perturbações que merecessem atenção mais aprofundada.

Desta forma optou-se por controlar, através de cartas de controlo, a qualidade da previsão, ou mais especificamente, do desvio da previsão face ao valor real, valor que se deseja o mais próximo de zero quanto possível. A maior dificuldade neste cenário será a grande variabilidade no comportamento das séries da previsão e do valor real, dado que são ambos resultados de processos naturais. Desta forma, optou-se neste estudo por modelar a série dos desvios da previsão (isto é, da diferença entre valor previsto da condutividade no reservatório e o valor real observado a cada instante) através dos modelos ARIMA, obtendo assim um termo “erro da série modelada” que será normalmente distribuído, e será esse termo que irá ser controlado através das cartas de controlo.

O primeiro passo será determinar exactamente qual a série dos desvios a modelar. Para esta monitorização da qualidade da previsão serão utilizados os dados obtidos através do modelo temporal com observações reais.

Dado que o modelo apresentado contém um intervalo inicial de dados onde o mesmo ainda não representa fielmente o que se passa na realidade, esses dados serão descartados. Ou seja, as primeiras 123 observações não serão incluídas na modelação ARIMA da série. O gráfico da Figura 5.14 representa a série dos desvios da qual se pretende ter uma modelação ARIMA.

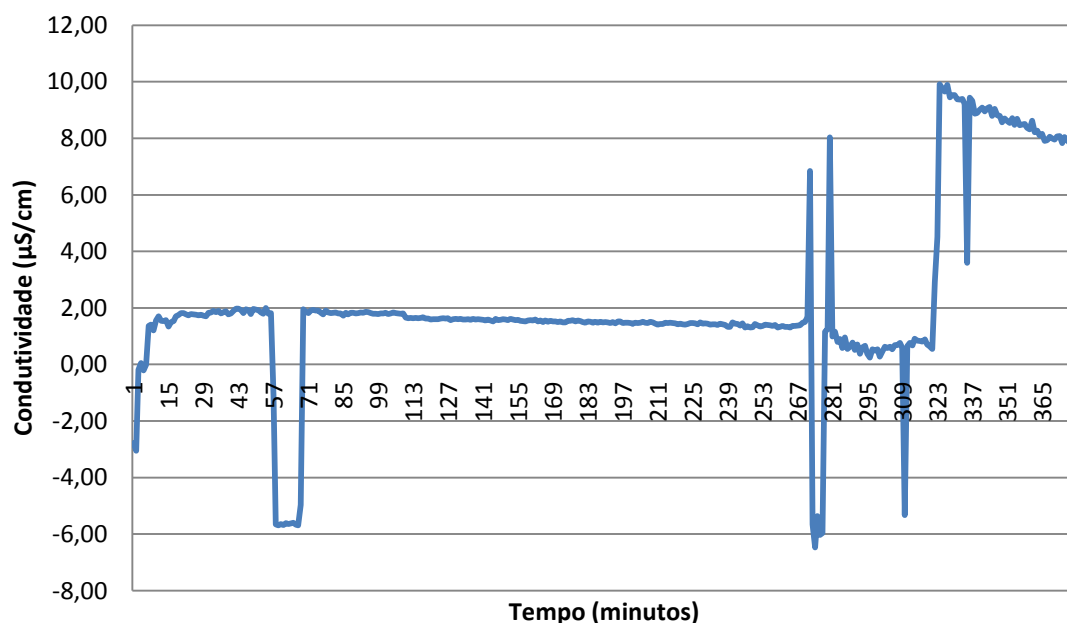


Figura 5.14 - Série dos desvios da previsão a modelar

Esta série constituída pelos restantes valores apresenta autocorrelação, como se pode verificar nos gráficos da ACF e da PACF, presentes na Figura 5.15 e Figura 5.16, respectivamente.

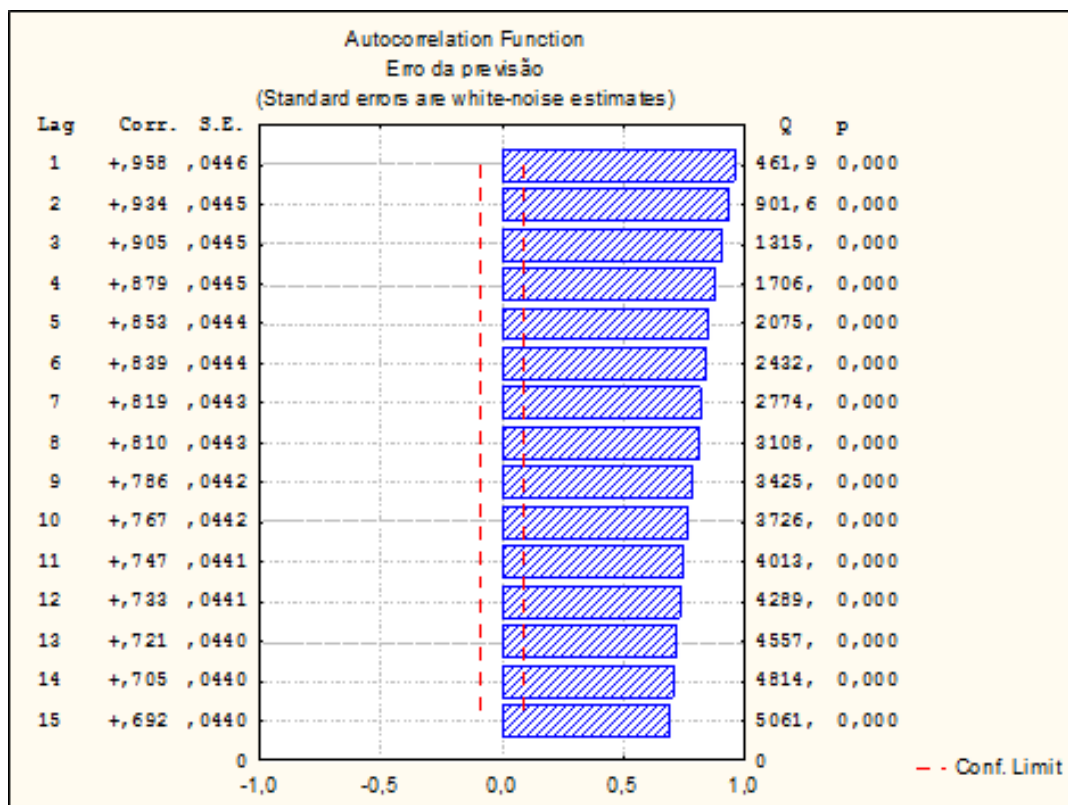


Figura 5.15 - Correlograma da série dos desvios da previsão

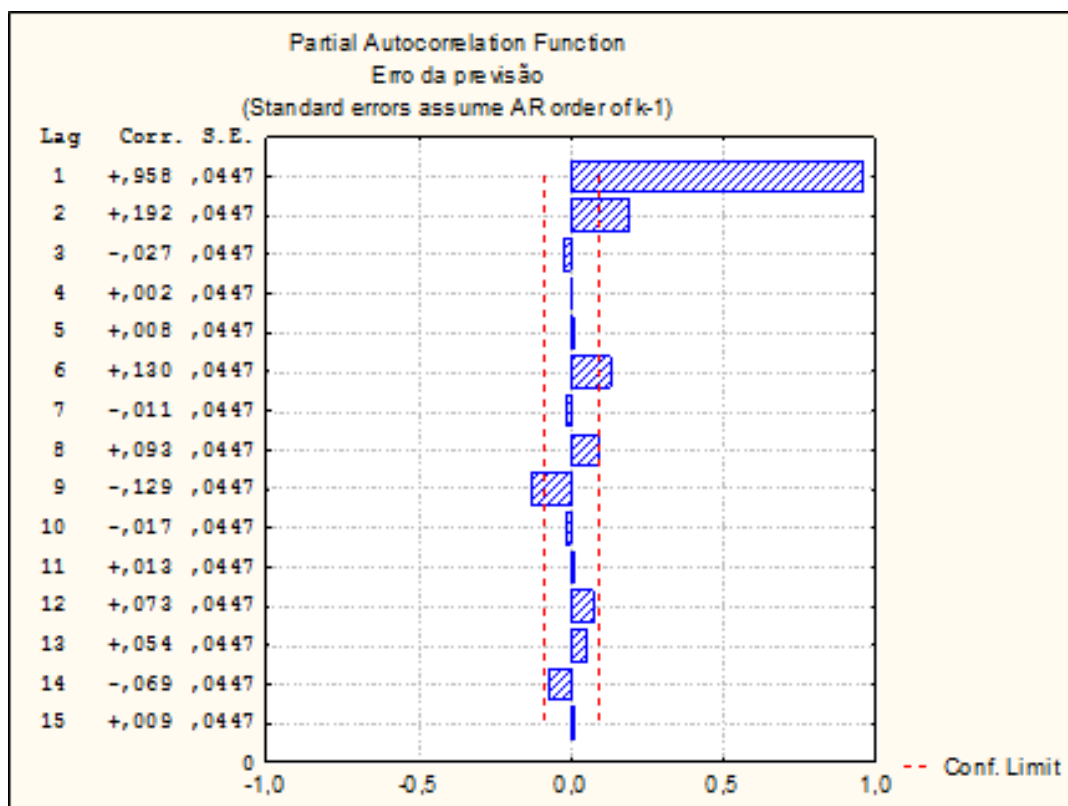


Figura 5.16- PACF da série dos desvios da previsão

Como se pode verificar dos gráficos da ACF e PACF, e de acordo com a Tabela 3.1, pode-se afirmar que se está perante uma série cujo modelo ARIMA mais adequado será o auto-

regressivo de 2º ordem. Dado ainda que se pode considerar ainda que o diferencial da série dos desvios da previsão apresentam comportamento de ruído branco, como se pode verificar nos gráficos do ACF e PACF do diferencial desta série, presentes no anexo III, respectivamente), pode-se afirmar que o modelo mais adequado para modelar esta série será o modelo ARIMA(2;1;0), ou auto-regressivo de 2º ordem com um diferencial de 1º ordem, sem componente *moving average*.

Através do *software statistica* é possível modelar a série com o modelo ARIMA acima referido, obtendo os parâmetros do modelo presentes na Tabela 5.10.

Tabela 5.10 - Parâmetros do modelo ARIMA da série dos desvios

Parâmetro	ϕ_0	ϕ_1	ϕ_2	SS	MS
Valor	0.028175	-0.236582	-0.002909	563.66	1.5112
Desvio padrão	0.02652	0.04468	0.04468	-	-

Partindo das equações (3.8) e (3.11), é possível demonstrar que a expressão geral de um modelo ARIMA (2;1;0) é a que se encontra na equação (5.1).

$$Y_t = Y_{t-1} + \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + (e_t - e_{t-1}) \quad (5.1)$$

A partir da expressão anterior, é possível determinar uma nova expressão para calcular, para cada instante, o valor do erro do termo ARIMA, termo esse que, caso siga uma distribuição normal, será monitorizado através de cartas de controlo. Essa expressão encontra-se na equação (5.2).

$$e_t = Y_t - Y_{t-1} - \phi_1(Y_{t-1} - Y_{t-2}) - \phi_2(Y_{t-2} - Y_{t-3}) + e_{t-1} \quad (5.2)$$

A partir da equação anterior, e considerando o erro do instante zero como sendo nulo, obtém-se a série do termo do erro que se apresenta no gráfico da Figura 5.17.

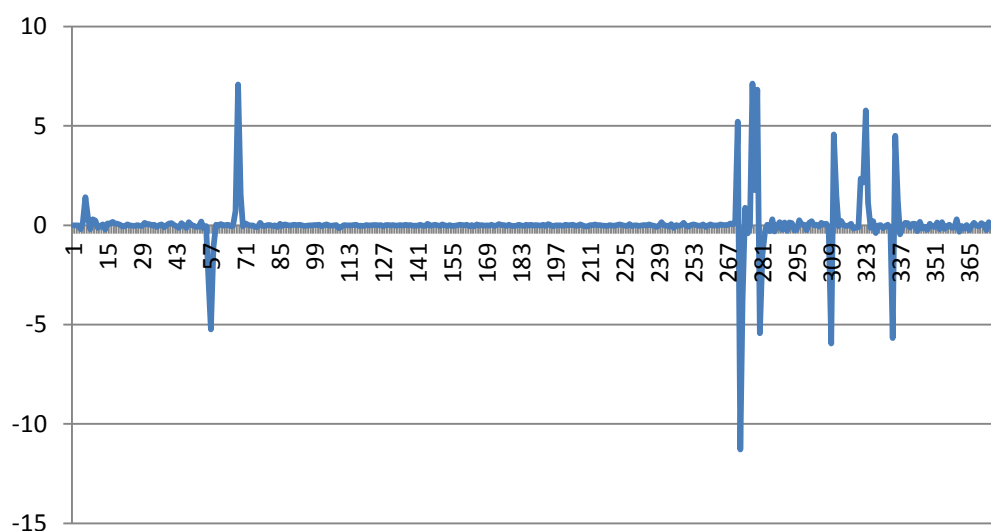


Figura 5.17 - Série do termo do erro do modelo ARIMA (2; 1; 0)

A série apresentada na figura anterior representa o resultado da modelação da série dos desvios da previsão. Aplicando um modelo ARIMA (2; 1; 0) a essa série de dados, é possível obter para

cada instante o valor do termo erro da modelação ARIMA, dado que a série Y_t é conhecida. Esse termo erro da modelação ARIMA é o que irá ser controlado através das cartas de controlo.

Para eliminar eventuais influências da premissa da nulidade do erro no instante inicial e compensando as falhas iniciais do modelo da previsão, da série anterior do gráfico anterior descartam-se as primeiras 70 observações, e às 100 seguintes verifica-se a normalidade, de modo a que seja possível aplicar a fase I das cartas tradicionais de Shewhart. Na Figura 5.18 pode-se encontrar o resultado obtido pelo estatístico do ajustamento à distribuição normal das 100 observações retiradas da série do gráfico anterior.

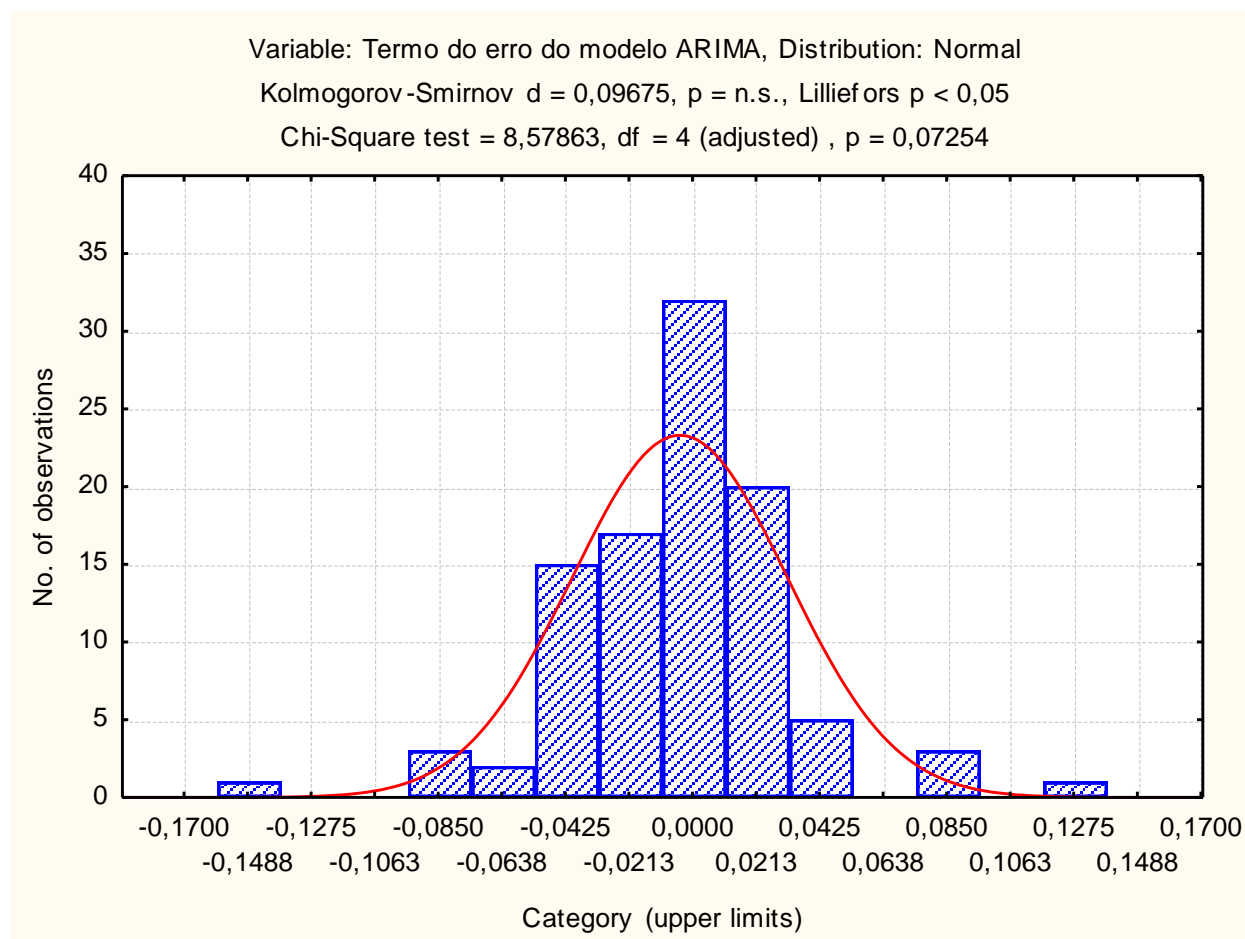


Figura 5.18 - Ajustamento à distribuição normal de 100 observações do termo do erro do modelo ARIMA

Na Tabela 5.11 apresentam-se os resultados obtidos do teste à normalidade das 100 observações do termo do erro do modelo ARIMA através do *software* statistica.

Tabela 5.11 - Resultados obtidos do teste à normalidade pelo estatística

Teste KS		Teste do Qui-Quadrado		
Valor obtido	Valor crítico tabelado	Valor obtido	Graus de liberdade	Valor crítico tabelado
0.09675	0.136	8.57863	4	9.488

Analisando a tabela anterior, pode-se verificar que, para ambos os testes, o valor crítico da estatística (retirados de tabelas presentes no anexo II, para o teste KS e Qui-Quadrado) é sempre superior ao valor obtido do *software* estatística para a estatística de teste, logo, para um nível de confiança de 95%, não se rejeita a hipótese nula, podendo-se então tratar esta população como sendo normalmente distribuída. A fase I do controle estatístico do processo será então aplicada, através das cartas de controlo tradicionais de Shewhart, nomeadamente as cartas X (para a média) e MR (para controlo da dispersão).

5.6.1 Fase I do controlo da previsão

Dado que existem observações suficientes para um controlo da previsão da condutividade, não será necessário recorrer às cartas short run para a monitorização da mesma.

Para aplicar as cartas tradicionais de Shewhart à série cuja distribuição se provou normal através dos testes do Qui-Quadrado e KS, e dado que os dados são referentes a observações individuais, utilizam-se as cartas X e MR, explicadas na subsecção 3.5.6.

Os resultados obtidos a partir da primeira carta X e carta MR encontram-se na Figura 5.19 e Figura 5.20, respectivamente.

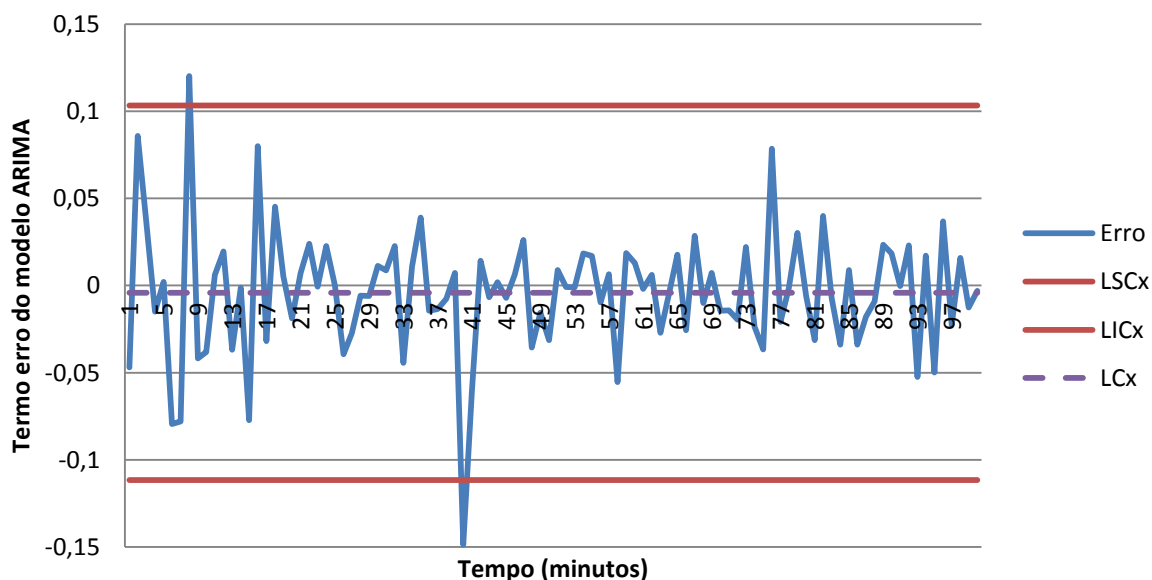


Figura 5.19 - Primeira iteração da carta X para o controlo do termo erro do modelo ARIMA

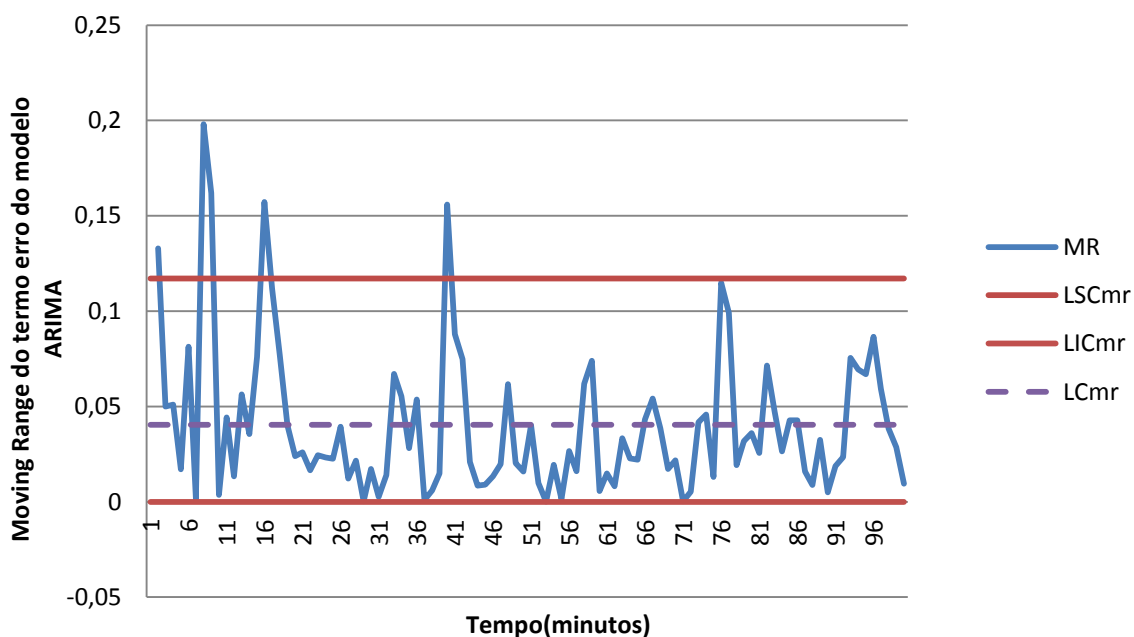


Figura 5.20 - Primeira iteração da carta MR para o controlo do termo erro do modelo ARIMA

Como se pode observar pela carta de controlo de observações individuais há 2 amostras que saem fora dos limites de controlo, e além destas duas há mais uma da carta MR que também ultrapassa o limite superior de controlo.

Estas três observações, mais especificamente, a observação 8, 16 e 40, são retiradas do estudo e os seus valores não terão contribuição para a determinação dos parâmetros do processo na fase I. Em substituição destas 3 observações, as 3 observações posteriores entram no conjunto da amostra da população a estudar para a fase I. As cartas X e MR que daí resultam podem-se encontrar no anexo III.

Na segunda iteração do processo de aplicação detectaram-se novamente observações fora dos limites de controlo, mais especificamente as observações 2 e 73, que ultrapassam o limite superior de controlo. O procedimento anterior é repetido, estas observações são retiradas e outras duas são incluídas à série de dados.

O resultado da terceira iteração da aplicação das cartas X e MR resulta nas cartas que se podem encontrar na Figura 5.21 e Figura 5.22.

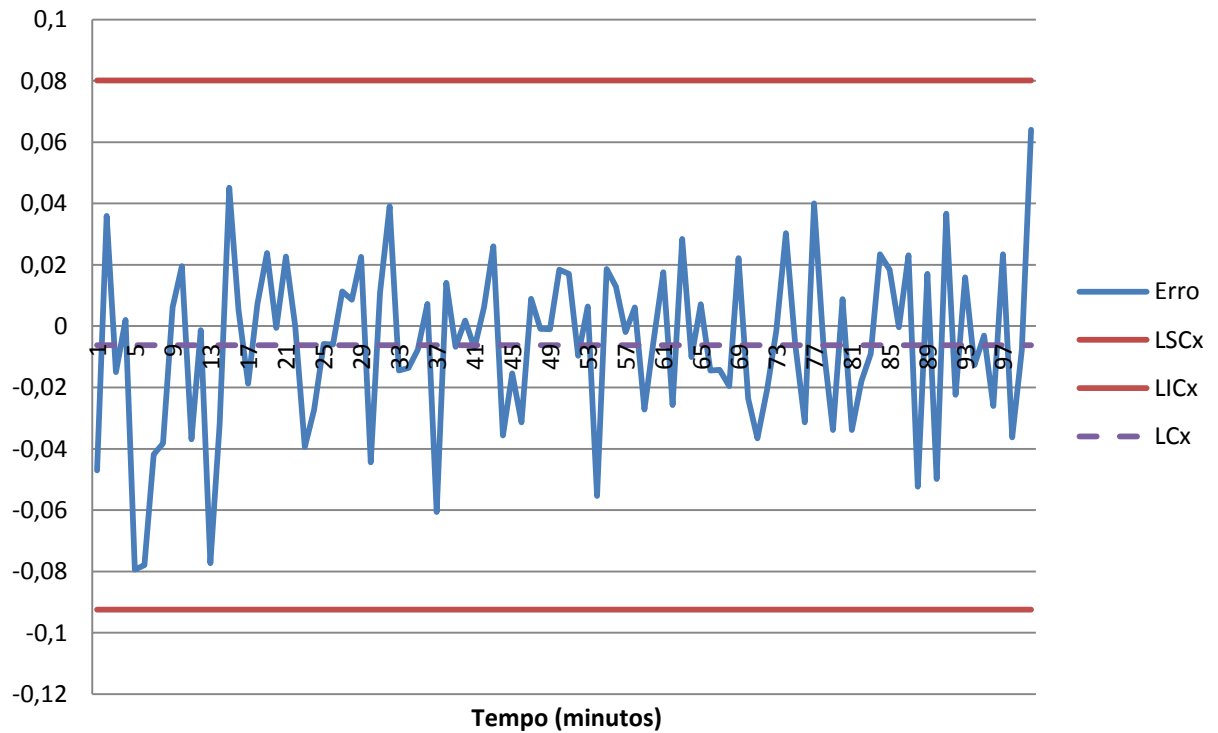


Figura 5.21 - Terceira iteração da carta X para o controle do termo erro do modelo ARIMA

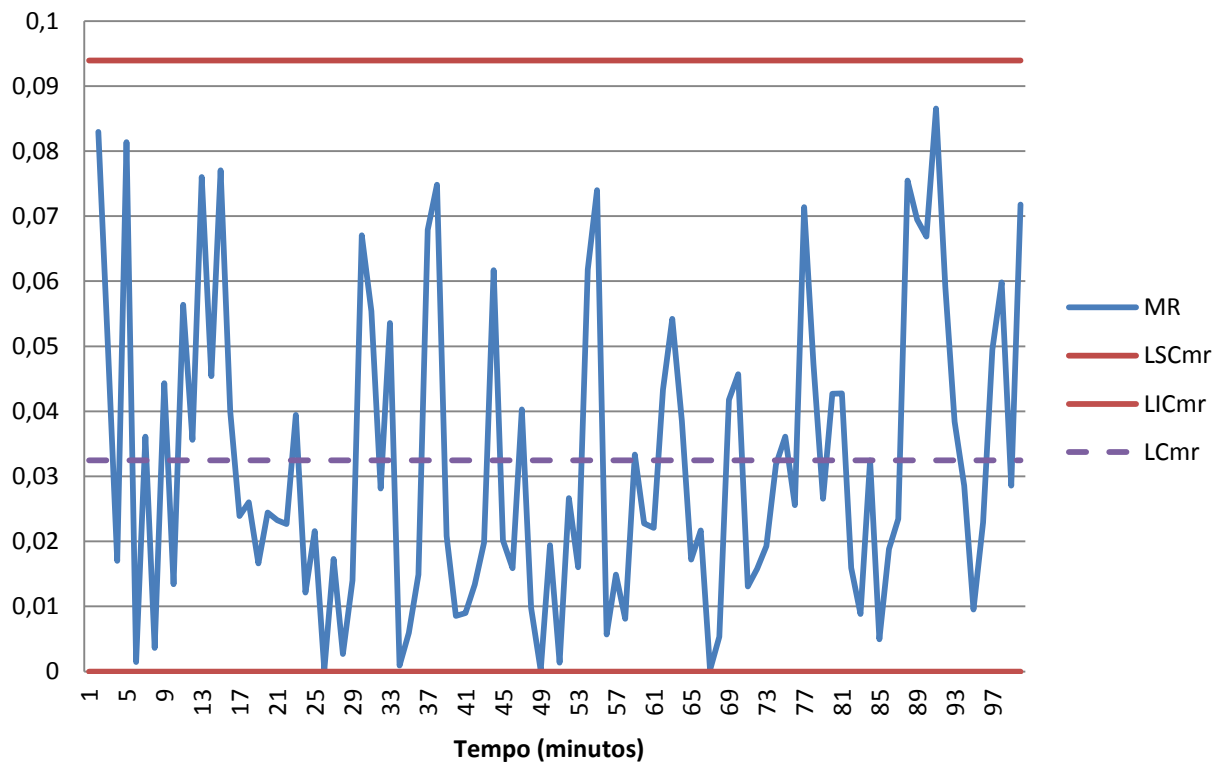


Figura 5.22 - Terceira iteração da carta MR para o controle do termo erro do modelo ARIMA

As cartas anteriores apresentam todos os pontos dentro dos limites de controle. No entanto, na carta X pode-se detectar a presença de uma das regras ISO, mencionadas na subsecção 3.5.2, mais especificamente a regra nº 4 (Catorze pontos consecutivos decrescendo e crescendo alternadamente). Dado que a eliminação dessas observações daria origem a um extenso

procedimento de remoção de observações, optou-se por se ignorar este facto, e considerar-se que o processo está sob controlo estatístico. Por se tratar de um processo natural, considera-se que a violação desta regra possa ocorrer sem constituir uma situação anormal. Dessa forma foi possível estimar os parâmetros da série dos erros do modelo ARIMA, que se resumem na Tabela 5.12.

Tabela 5.12 - Parâmetros estimados para a série dos erros do modelo ARIMA

Parâmetros do processo:	
Valor esperado ($\hat{\mu}$)	-0.00618
Variância ($\hat{\sigma}^2$)	0.028756

5.6.2 Fase II do controlo da previsão

A partir dos dados apresentados na tabela anterior é possível partir para a fase II do controlo estatístico do processo, com os restantes dados da série de erros do modelo ARIMA.

Para tal, utilizam-se as cartas e MCEWMA e EWMA de resíduos, com o intuito de comparar a performance de ambas para o controlo estatístico deste tipo de dados.

De forma a construir estas cartas, o primeiro passo foi criar uma série de dados a partir da expressão obtida a partir do modelo ARIMA. Esta série teórica representará a série de erros de uma previsão eficaz, e como tal terá todos os seus pontos dentro dos limites de controlo das cartas MCEWMA e EWMA. A partir dos parâmetros presentes na Tabela 5.10 e da expressão (5.2), e introduzindo um erro em cada instante com distribuição normal cuja média e desvio-padrão são retirados a partir da Tabela 5.12 pode-se construir uma série da previsão dos termos do erro do modelo ARIMA que represente uma previsão da condutividade eficaz. A partir desta série constroem-se as cartas MCEWMA e EWMA de resíduos, onde posteriormente serão inseridos os dados reais da série dos termos do erro do modelo ARIMA (ou dados frescos), que permitirão o controlo em tempo real da qualidade da previsão do modelo temporal do *software matlab*.

Para a construção da carta MCEWMA, o primeiro passo será a definição do parâmetro λ , a partir do erro quadrático mínimo. O valor deste parâmetro será determinado a partir do ponto mínimo do gráfico resultante da aplicação deste método, representado na Figura 5.23.

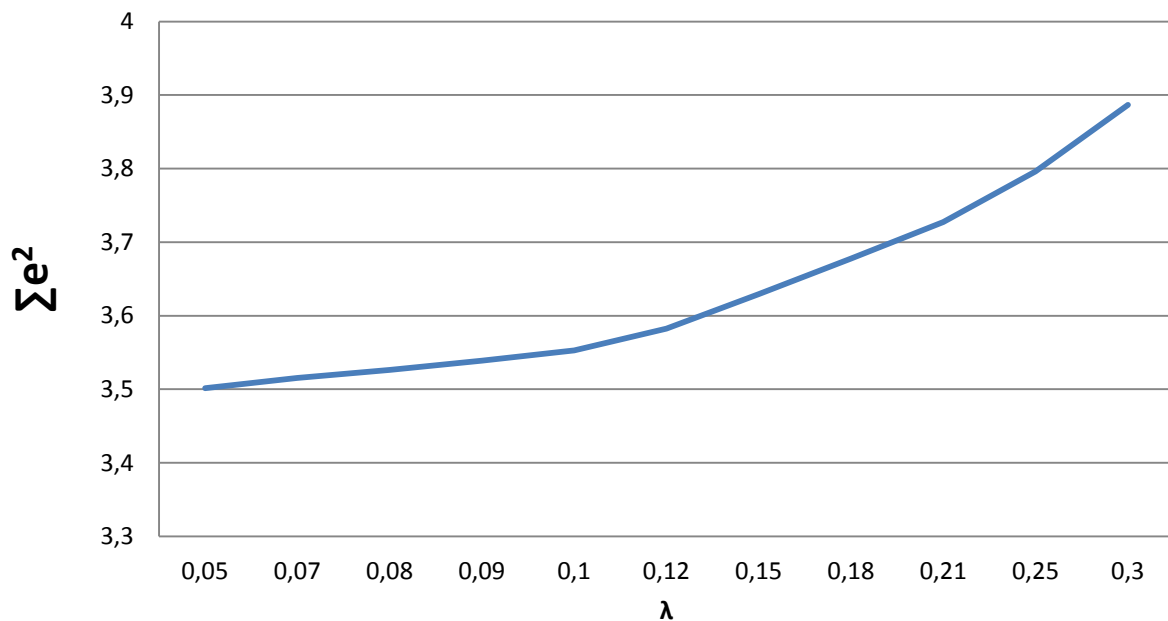


Figura 5.23 - Valor do somatório do erro ao quadrado em função de λ

Como se pode observar na figura anterior, o valor de λ que minimiza o erro quadrático é de 0,05, sendo portanto este valor o utilizado nas cartas MCEWMA apresentadas de seguida.

A partir das equações presentes na subsecção 3.7.3 e do parâmetro λ definido é possível construir a carta MCEWMA presente na Figura 5.24, com os dados previstos a partir do modelo ARIMA, representando o controlo de uma previsão eficaz.

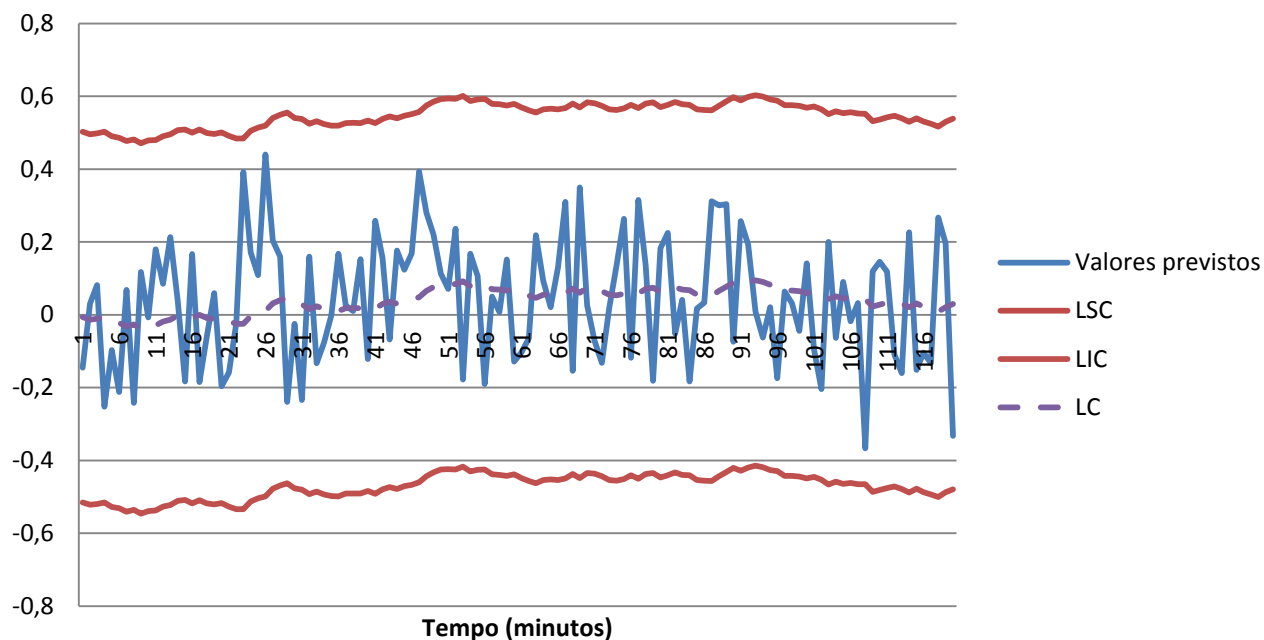


Figura 5.24 - Carta MCEWMA para os valores previstos do modelo ARIMA

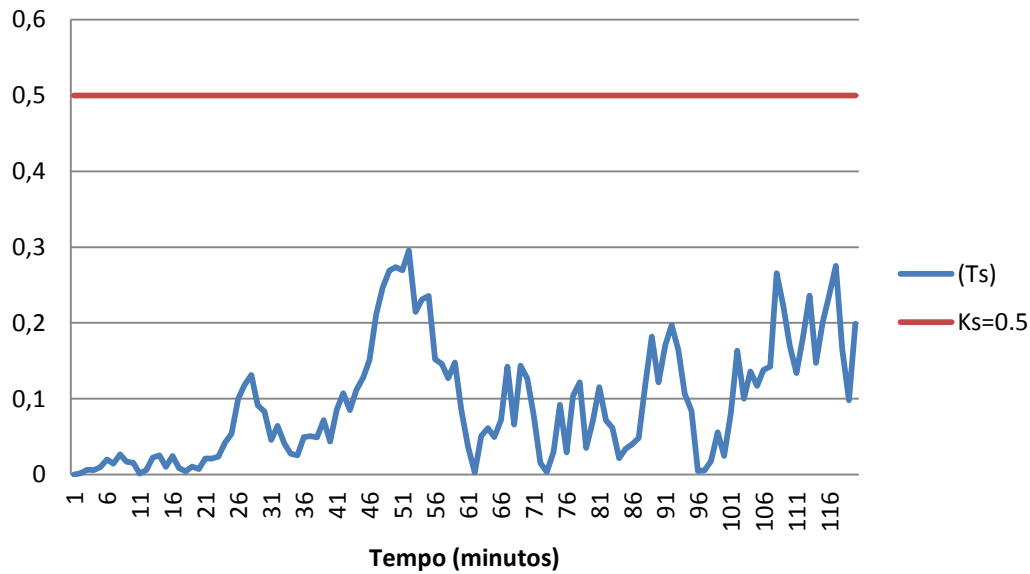


Figura 5.25 - Gráfico do índice T_s para os dados previstos do termo erro do modelo ARIMA

Como seria de esperar, para uma previsão eficaz os valores da estatística E da carta MCEWMA estão todos dentro dos limites de controlo, o que indica uma previsão eficaz. O passo seguinte será inserir nesta carta os dados frescos, ou seja, os dados reais observados do termo erro do modelo ARIMA, e observar o comportamento da estatística E ao longo do tempo. Na Figura 5.26 apresenta-se o resultado desse mesmo controlo.

No que toca ao índice T_s , na figura 5.25 pode-se encontrar o gráfico dos valores desse mesmo índice ao longo do tempo e do valor crítico considerado.

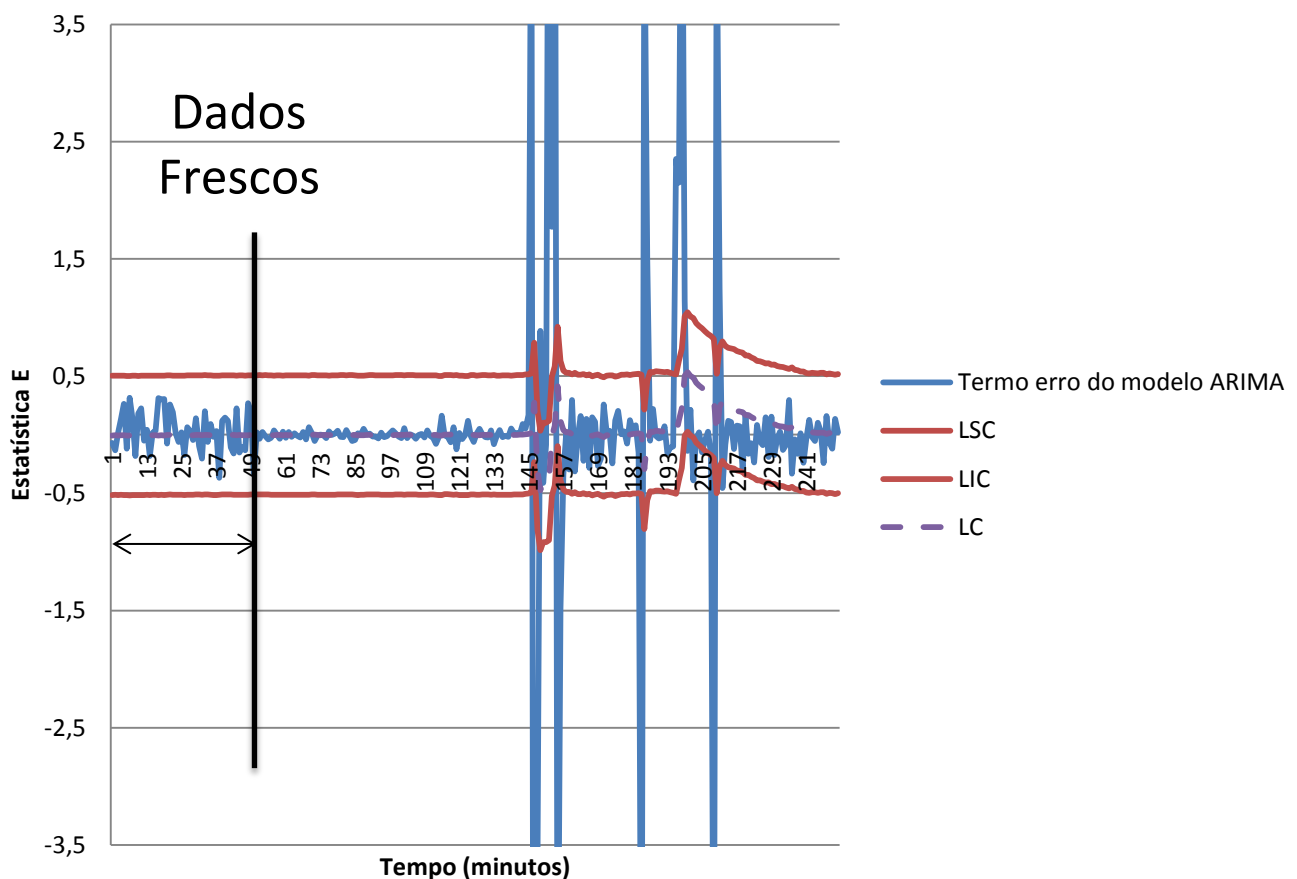


Figura 5.26 - Carta MCEWMA com os dados observados do termo erro do modelo ARIMA

A carta apresentada na figura anterior representa a junção dos dados frescos às últimas 50 observações previstas apresentadas na carta da Figura 5.24, separadas pela semi-recta vertical de cor preta. Pode-se verificar que a série dos termos do erro do modelo ARIMA sempre se encontra sob controlo até ao momento em que a previsão se afasta mais do valor real, onde esta carta apresenta a maior variação dos limites e onde surgem os pontos fora dos limites de controlo. A mesma observação se retira do gráfico do índice T_s com os dados frescos inseridos após os dados previstos, que se pode encontrar na Figura 5.27.

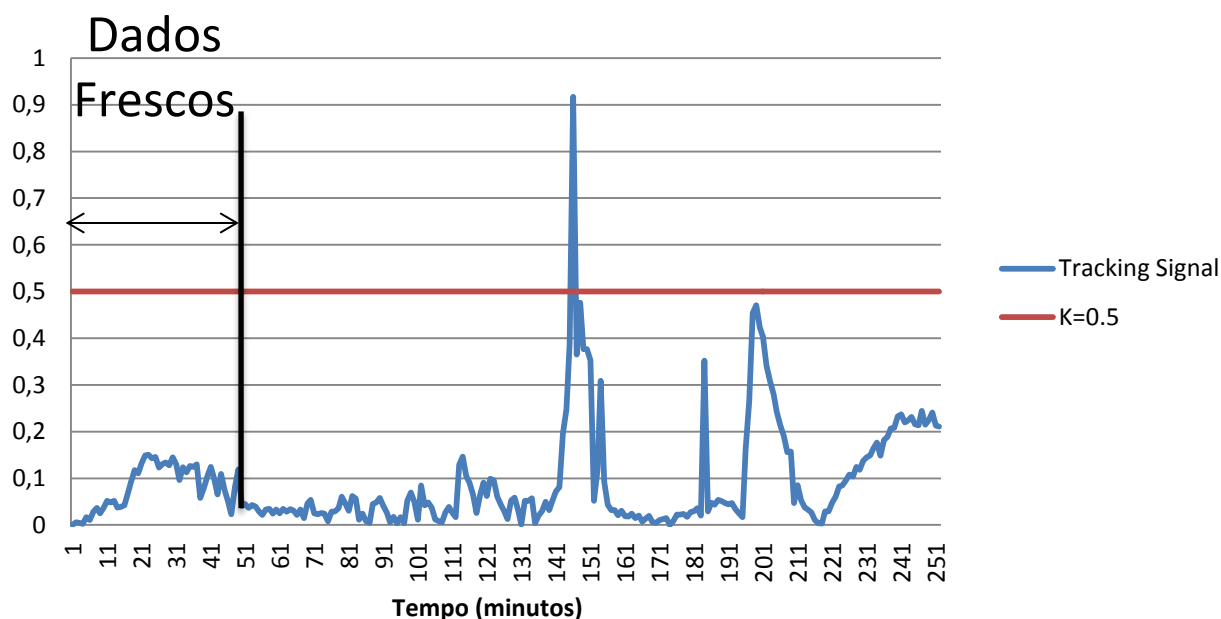


Figura 5.27 - Gráfico do índice Ts para os dados frescos da série do termo erro do modelo ARIMA

A mesma metodologia foi aplicada para o controlo da fase II com as cartas EWMA de resíduos, explicadas na subsecção 3.7.2 onde se apresenta na Figura 5.28 a carta para os dados previstos para o termo do erro do modelo ARIMA, tomando como desvio da média a controlar de 1.5σ e um valor de ARL_{EC} de 370.

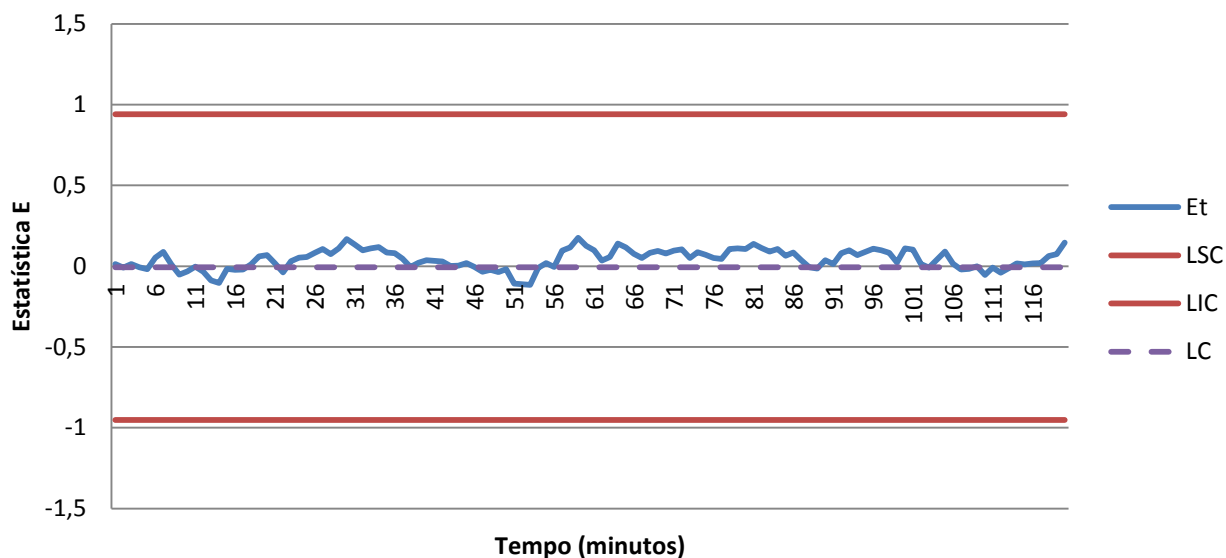


Figura 5.28 - Carta EWMA para os valores previstos do termo erro do modelo ARIMA

Novamente se verificam todos os pontos dentro dos limites de controlo, indicando como se deverá comportar na carta EWMA uma série de previsão eficaz. Na Figura 5.29 apresenta-se a carta EWMA com os dados observados do termo erro do modelo ARIMA, antecidos pelos últimos 50 termos previstos pelo mesmo modelo, separados pela semi-recta a preto.

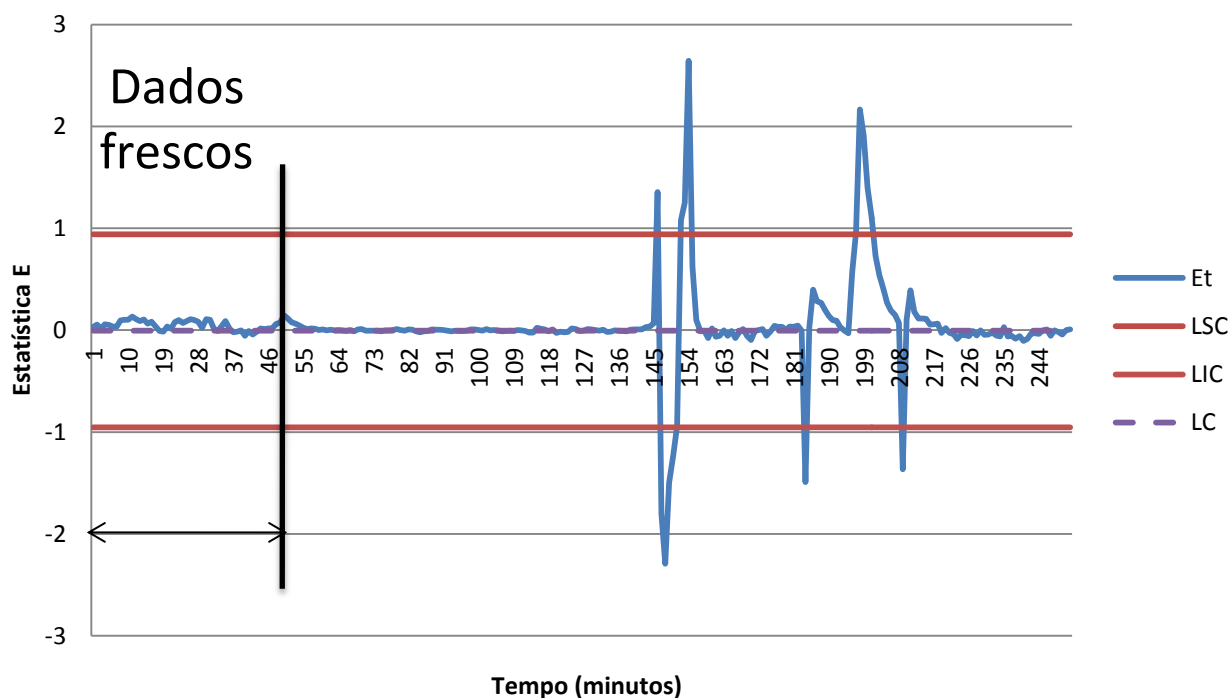


Figura 5.29 - Carta EWMA com os dados observados do termo erro do modelo ARIMA

Pode-se observar que, à semelhança da carta MCEWMA, a carta EWMA mostra que a previsão se encontra sob controlo e a partir de um dado instante no tempo começa a desviar-se do valor real, apresentando os pontos fora dos limites de controlo. Nesse momento será exigida uma investigação para determinar as causas do desvio entre a previsão da condutividade da água no reservatório de saída e o seu valor real.

Este valor pode dever-se a eventuais infiltrações de agentes poluentes na rede de distribuição, ou desvios de volumes de água para subestações secundárias não contabilizados pela telegestão.

5.7 Capacidade do processo

A nível de especificações legais, a água distribuída nesta rede pelo SMAS de Almada encontra-se bastante abaixo dos valores indicados como guia para a água para consumo humano. Valor guia esse, como se citou anteriormente, encontra-se nos 2500 $\mu\text{S}/\text{cm}$ a 20 °C. Como tal considera-se redundante um estudo dos índices de capacidade, tanto para as cartas de controlo tradicionais como para as especiais. De qualquer forma, caso seja necessário aplicar esse estudo noutras redes com valores mais próximos dos mencionados na lei nacional, podem ser aplicadas as técnicas referidas nas subsecções 3.5.3 (para as cartas tradicionais) 3.6.4 (para as cartas Q).

Capítulo 6 – Conclusões e recomendações

6.1 Conclusões

O propósito da existência deste estudo foi o de criar um modelo de previsão da condutividade em reservatórios da rede do SMAS de Almada, que permitisse o controlo dos níveis dessa característica e que incluísse a variável tempo no mesmo, de modo que o modelo se comportasse de uma forma dinâmica.

A característica escolhida para monitorizar a qualidade da água foi a condutividade, pois é um indicador da presença de poluentes indesejados que é relativamente económico e rápido de analisar.

Para esse efeito foi estudada uma sub-rede de abastecimento de água dos SMAS, que foi caracterizada na sua totalidade, a nível de entradas (quer de outros reservatórios ou de captações subterrâneas), comprimentos, áreas internas e interligações da rede de tubagens e saída (reservatório de saída).

O sistema de telegestão a ser implementado no SMAS de Almada permitiu a recolha de dados com qualquer periodicidade desejada, o que permite ter informações actualizadas sobre os dados do sistema em pontos estratégicos da rede, apesar do sistema ainda apresentar algumas falhas nos dados, causadas pelo facto de o sistema ainda se encontrar numa fase relativamente embrionária no que toca às suas capacidades de controlo de variáveis.

Estas falhas nas séries de dados puderam ser corrigidas, usando para isso o *software statistica*, que tem a capacidade de, partindo dos dados reais disponíveis, estimar os dados que a série tem em falta. Apesar de não conseguir reproduzir fielmente o comportamento das séries reais, devido à sua variabilidade natural, consegue ainda assim ser um bom estimador para inserir no modelo de previsão.

Os modelos propostos que incluem a variável tempo (modelo temporal e modelo híbrido) mostraram uma boa capacidade de replicar o comportamento da série real de dados, com um desvio médio absoluto menor do que a própria incerteza de medição do aparelho utilizado para medir a condutividade da água no SMAS de Almada, mas isto só se verifica para séries de dados onde a periodicidade de recolha de dados é mais curta.

Para maiores intervalos de tempo entre recolha de dados um modelo mais indicado será o modelo não-temporal, pois é aquele que se mostrou mais capaz de reproduzir o comportamento da série, embora com erros de previsão superiores aos dos modelos temporais com intervalos de tempo entre recolhas de dados mais reduzidos.

Isto implica que, com a capacidade mostrada pelo sistema de telegestão na recolha de dados, os modelos temporais provam ter a capacidade de prever adequadamente a característica da condutividade num reservatório, tendo como input os dados relativamente às suas entradas de água e configuração da rede de condutas.

No entanto existem alguns factores que influenciam o erro e que devem ser tidos em conta:

- Podem existir, em determinados troços e em determinados instantes de tempo, ligações de uma rede em estudo para outras subestações e reservatórios que desviam alguma da água que se julgava ter como destino o reservatório de saída do sistema. Estes desvios não se encontram contabilizados na série de dados, e resulta que na realidade, numa entrada específica, esteja a entrar menos água do que se utiliza no modelo para fins de cálculo. Esta situação ocorre, na rede em estudo, na entrada E1;

- A mistura da água com condutividades diferentes não é instantânea, existindo uma heterogeneidade, que depende também da geometria do reservatório de saída e da forma como as entradas e saídas de água se localizam no mesmo;
- Pode haver mistura de águas com condutividades diferentes dentro de uma conduta, especialmente em situações com alta variabilidade de caudal, e esta situação não se encontra englobada no modelo;
- A periodicidade com que é controlada a água das captações subterrâneas pode fazer com que os parâmetros da água daí retirada num dado instante não correspondam aos esperados pelo modelo;
- Existência de subestações secundárias que retiram água das condutas sem que esse volume seja registado pela telegestão, gerando diferenças entre o volume de água real das condutas ou reservatórios e o volume esperado pelo modelo.

Duas características importantes do modelo é que o mesmo pode ser aplicado em diversos tipos de geometria de reservatório, mudando apenas a forma como o volume de água existente no interior do mesmo em função do nível da água lá registado. A outra é que o mesmo será aplicável para outras características da qualidade da água, para além da condutividade. Existindo uma expressão matemática que defina os valores desse parâmetro em função dos valores iniciais das águas que entram na mistura, a mesma pode ser aplicada no modelo, do qual se espera previsões dessa característica com a mesma qualidade do que as previsões da condutividade.

Os modelos temporais apresentados devem, agora que validados, ser aplicados em cadeia, pois existem diversas situações onde um reservatório final de um subsistema é, ele próprio, uma entrada de um outro subsistema na rede dos SMAS de Almada. Implementando um sistema capaz de aplicar estes modelos temporais em cadeia, é possível ter previsões num reservatório, que serão utilizadas, em conjunto com outras previsões e/ou dados reais de uma captação subterrânea, para fazer nova previsão para um outro subsistema.

Os factos anteriores fazem com que o modelo seja uma ferramenta poderosa para aplicar em situações de produção de água para consumo humano e seu controlo em redes de distribuição.

Através da realização deste trabalho concluiu-se ainda que as cartas de controlo *short run*, mais especificamente as cartas Q, podem constituir um método eficaz para controlo da água retirada de captações subterrâneas, permitindo o controlo e actualização dos parâmetros dessa mesma água para a sua posterior inserção no modelo de previsão, além de identificar eventuais contaminações nas captações subterrâneas. Estas cartas apresentaram um funcionamento satisfatório porque os valores recolhidos nas captações apresentam um comportamento por patamares, em parte devido à periodicidade de recolha dos valores da condutividade.

No entanto estas cartas não se mostram eficazes para todos os tipos de inputs do modelo. Para entradas na rede cujos parâmetros tenham uma variabilidade extrema, como é o caso dos reservatórios, outro tipo de metodologia é aconselhável, como por exemplo o estudo por patamares realizado nesta dissertação. Este tipo de entradas de água na rede merece um controlo muito mais apertado, como se pode concluir a partir da análise de sensibilidade realizada neste estudo, onde flutuações no reservatório que servia como entrada de água na rede tinham maiores repercussões na qualidade da previsão final.

A partir da análise de sensibilidade determinou-se que pequenas flutuações nas captações subterrâneas não têm um efeito significativo na qualidade da previsão realizada pelo modelo, como tal o controlo às mesmas realizado actualmente será suficiente para suportar o modelo de previsão proposto, ao invés do que acontece nas entradas de água a partir de reservatórios, onde a maior variabilidade implicará um controlo mais apertado.

No que toca à monitorização da qualidade da previsão, pode-se concluir que no caso apresentado neste estudo tanto as cartas EWMA como MCEWMA apresentam performances semelhantes, pois foram capazes de detectar ao mesmo tempo quando a previsão da condutividade do reservatório de saída deixou de ser eficaz, sendo que a carta MCEWMA apresenta as vantagens da aplicabilidade a séries de dados não estacionárias e de ser mais sensível às variações da série a controlar, embora seja de construção mais complexa.

Ambas as cartas detectaram o ponto onde a previsão se afasta do valor real pelos motivos anteriormente referidos, o que mostra que qualquer uma delas pode ser uma ferramenta eficaz no controlo da previsão da condutividade da água numa rede.

Realizando este estudo foi possível concluir que, garantindo o controlo dos inputs de água, realizando uma previsão do valor da saída e controlando essa mesma previsão é possível um controlo eficaz da rede de recolha e distribuição de água numa rede de condutas. Desta forma será possível definir níveis de alerta dinâmicos que permitam detectar eventuais contaminações poucos momentos após o seu início, permitindo agir atempadamente e evitar consequências graves ao sistema de produção de água para consumo humano.

6.2 Sugestões para trabalhos futuros

O presente estudo centrou-se no desenvolvimento de um modelo capaz de prever a condutividade da água numa rede de condutas e reservatórios de entrada e saída da água, onde o tempo é uma variável crucial. Desenvolvido e validado o mesmo, o passo seguinte será a sua implementação.

Para tal, aconselha-se a utilização de um *software* capaz de receber dados em tempo real e processá-los, utilizando o modelo proposto, tendo a capacidade de guardar as séries de dados recolhidas e utilizadas como input no modelo, e descartar os dados que já não serão necessários para a aplicação do modelo, dado que o mesmo não precisa de um histórico de dados demasiado grande.

Sugere-se ainda o estudo da aplicabilidade dos modelos apresentados para outros parâmetros da qualidade da água, sendo necessários estudos laboratoriais para definir, num modelo mais completo, a lei matemática que rege o resultado da mistura de água com parâmetros diferentes num outro parâmetro que não a condutividade, para uma água com determinadas características.

Determinadas estas regras a serem inseridas no modelo, o mesmo pode ser implementado em cadeia no SMAS de Almada, permitindo a realização da previsão dos parâmetros da qualidade da água de forma dinâmica e em tempo real, com a capacidade de actuar sobre um eventual foco de discrepância entre o valor previsto e valor real da forma que se considere mais indicada.

Questões relacionadas com a dinâmica de fluidos podem ainda ser aprofundadas neste ambiente: sugerem-se estudos sobre a influência de variações bruscas de caudal na velocidade com que volumes de água passam pela rede, e no modo como a geometria dos reservatórios e suas ligações influenciam a mistura de águas com características distintas.

Aplicando o modelo apresentado em cadeia, e a toda a rede de distribuição do SMAS será possível ter uma previsão e controlo eficaz e em tempo real das características da água que será fornecida aos consumidores, melhorando o serviço prestado por esta entidade pública.

Bibliografia

- Aizenchtadt, E., Ingman, D., & Friedler, E. (2008). Quality control of wastewater treatment: A new approach. *European Journal of Operational Research*, 189(2), 445–458.
- Box, G., Jenkins, G., & Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control* (3rd ed.). Prentice-Hall.
- Brockwell, P. J., & Davis, R. A. (2002). *Introduction to Time Series and Forecasting* (2nd ed.). Springer.
- Chakraborti, S. (2007). Run length, average run length and false alarm rate of shewhart x-bar chart: exact derivations by conditioning. *Communications in Statistics - Simulation and Computation*, 29(1), 61–81. doi:10.1080/03610910008813602
- Cliff, J. K. (1983). Quality control in water analysis. *Environmental Science & Technology*, 17(4), 174A–181A. doi:10.1021/es00110a715
- Cun, C., & Vilagines, R. (1997). Time series analysis on chlorides, nitrates, ammonium and dissolved oxygen concentrations in the Seine river near Paris. *Science of The Total Environment*, 208(1-2), 59–69.
- Decreto-Lei n.º236/98 de 1 de Agosto. Diário da República n.º.176/98 - I Série A. Ministério do Ambiente. Lisboa.
- Decreto-Lei n.º306/07 de 27 de Agosto. Diário da República n.º.164/07 - I Série A. Ministério do Ambiente, do Ordenamento do Território e do Desenvolvimento Regional. Lisboa.
- Edzwalds, J. K. (Ed. lit.). (2010). *Water Quality & Treatment: A Handbook on Drinking Water* (6th ed.). American Water Works Association.
- European Charter on Water Resources. (n.d.). Acedido em 15/Maio/ 2013, de <http://unterm.un.org/DGAACS/unterm.nsf/8fa942046ff7601c85256983007ca4d8/3606b8495d1e10d68525783b0057ed3f?OpenDocument>
- Follador, F. A., Boas, M. A., Schoenhals, M., Hermes, E., & Rech, C. (2012). Tabular CUSUM control charts of chemical variables applied to the control of surface water quality. *Engenharia Agrícola*, 32(5), 951–960.

- Gustafson, H., & Behrman, A. S. (1939). Determination of Total Dissolved solids in Water by Electrical Conductivity. *Industrial & Engineering Chemistry Analytical Edition*, 11(7), 355–357. doi:10.1021/ac50135a001
- Hyndman, R., & Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Juran, J., & Godfrey, B. (1998). *Juran's Quality Handbook* (5th ed.). McGraw-Hill.
- Kaya, A. (2010). Statistical Modelling for Outlier Factors. *Ozean Journal of Applied Sciences*, 3(1), 185–194.
- Makridakis, S., Wheelwright, S., & Hyndman, R. (1997). *Forecasting Methods and Applications* (3rd ed.). Wiley.
- Modarress, B., Ansari, A., & Willis, G. (2000). Controlled production planning for Just-In-Time short-run suppliers. *International Journal of Production Research*, 38(5), 1163–1182. doi:10.1080/002075400189077
- Oakland, J. S. (2007). *Statistical Process Control* (6th ed.). Butterworth-Heinemann.
- Paterson, A. R. (1983). *A First Course in Fluid Dynamics*. Cambridge University Press.
- Pereira, Z. L., & Requeijo, J. G. (2012). *Qualidade: Planeamento e Controlo Estatístico de Processos* (2.^a edição). Fundação da Faculdade de Ciências e Tecnologias da Universidade Nova de Lisboa.
- Pires, A. R. (2007). *Qualidade: Sistemas de Gestão da Qualidade* (3.^a edição). Edições Sílabo.
- Ramos, P. M., Pereira, J. M. D., & Ribeiro, A. L. (2008). A Four-Terminal Water-Quality-Monitoring Conductivity Sensor. *IEEE Transactions on instrumentation and measurement*, 57(3), 577–583.
- Sachs, L. (1984). *Applied Statistics: A Handbook of Techniques* (2nd ed.). Springer.
- Silva, A. (2012). *Desenvolvimento de um Modelo Estatístico para o Controlo da Condutividade da Água dos SMAS de Almada*. Dissertação de Mestrado, Universidade Nova de Lisboa, Monte da Caparica.

SMAS Almada. (2013). *Mapa de Pessoal 2013*. Retirado de:

http://www.smasalmada.pt/portal/page/portal/front_end_portal_smas_almada/recursos/imagens/conteudos/Quem%20somos/mp_2013.pdf

SMAS Almada. (n.d.-a). História. Acedido em 12/Fev/ 2013, de

http://www.smasalmada.pt/portal/page/portal/front_end_portal_smas_almada/artigo?paramFolderId=273766*273595*114*pt*historia

SMAS Almada. (n.d.-b). Organização/organograma. Acedido em 12/Fev/ 2013, de

http://www.smasalmada.pt/portal/page/portal/front_end_portal_smas_almada/artigo?paramFolderId=273863*273595*114*pt*organograma

SMAS Almada. (n.d.-c). Indicadores de Gestão. Acedido em 12/Fev/ 2013, de

http://www.smasalmada.pt/portal/page/portal/front_end_portal_smas_almada/artigo?paramFolderId=273733*273595*114*pt*indicadores

SMAS Almada. (n.d.-d). *Opções do plano 2013*. Acedido em 12/Fev/2013, de

http://www.smasalmada.pt/portal/page/portal/front_end_portal_smas_almada/recursos/imagens/conteudos/Investimentos%20no%20Futuro/Op%E7%F5es%20do%20Plano%20e%20Or%E7amento/op_2013.pdf

SMAS Almada. (n.d.-e). Sistema Abastecimento Água. Acedido em 18/Fev/ 2013, de

http://www.smasalmada.pt/portal/page/portal/front_end_portal_smas_almada/artigo?paramFolderId=274315*274005*114*pt*sistema_abastecimento_agua

SMAS Almada. (n.d.-f). Telegestão. Acedido em 18/Fev/ 2013, de

http://www.smasalmada.pt/portal/page/portal/front_end_portal_smas_almada/artigo?paramFolderId=38196236*274005*114*pt*Telegest%E3o

SMAS Almada. (n.d.-g). Laboratório de Água de Consumo. Acedido em 18/Fev/ 2013, de

http://www.smasalmada.pt/portal/page/portal/front_end_portal_smas_almada/artigo?paramFolderId=274430*274005*114*pt*laboratorio_agua_consumo

SMAS Almada. (n.d.-h). Qualidade Água. Acedido em 18/Fev/ 2013, de

http://www.smasalmada.pt/portal/page/portal/front_end_portal_smas_almada/artigo?paramFolderId=274352*274005*114*pt*1_qualidade_agua

- Smeti, E. M., Koronakis, D. E., & Golfinopoulos, S. K. (2007). Control charts for the toxicity of finished water—Modeling the structure of toxicity. *Water Research*, 41(12), 2679–2689.
- Smeti, E. M., Thanasoulas, N. C., Kousouris, L. P., & Tzoumerkas, P. C. (2007). An approach for the application of statistical process control techniques for quality improvement of treated water. *Desalination*, 213(1-3), 273–281.
- Taguchi, G., Chowdhury, S., & Wu, Y. (2004). *Taguchi's Quality Engineering Handbook* (1st ed.). John Wiley & Sons, Inc.
- US EPA, O. (n.d.). Conductivity. Acedido em 05/Fev/ 2013, de <http://water.epa.gov/type/rsl/monitoring/vms59.cfm>
- Usman, A., & Kontagora, N. M. (2010). Statistical Process Control on Production: A Case Study of Some Basic Chemicals Used in Pure Water Production. *Pakistan Journal of Nutrition*, 9(4), 387–391.
- Zhou, W., Beck, B. F., Pettit, A. J., & Wang, J. (2006). Application of water quality control charts to spring monitoring in karst terranes. *Environmental Geology*, 53(6), 1311–1321.

Anexos

Anexo I – Tabelas estatísticas e de cartas de controle

$\nu \backslash \alpha$.995	.975	.9	.5	.1	.05	.025	.01	.005	.001	α / ν
1	0.000	0.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	10.828	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	13.816	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	16.266	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	18.467	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	20.515	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	22.458	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	24.322	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	26.124	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	27.877	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	29.588	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	31.264	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	32.910	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	34.528	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	36.123	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	37.697	15
16	5.142	6.908	9.312	15.338	23.542	26.296	28.845	32.000	34.267	39.252	16
17	5.697	7.564	10.085	16.338	24.769	27.587	30.191	33.409	35.718	40.790	17
18	6.265	8.231	10.865	17.338	25.989	28.869	31.526	34.805	37.156	42.312	18
19	6.844	8.907	11.651	18.338	27.204	30.144	32.852	36.191	38.582	43.820	19
20	7.434	9.591	12.443	19.337	28.412	31.410	34.170	37.566	39.997	45.315	20
21	8.034	10.283	13.240	20.337	29.615	32.670	35.479	38.932	41.401	46.797	21
22	8.643	10.982	14.042	21.337	30.813	33.924	36.781	40.289	42.796	48.268	22
23	9.260	11.688	14.848	22.337	32.007	35.172	38.076	41.638	44.181	49.728	23
24	9.886	12.401	15.659	23.337	33.196	36.415	39.364	42.980	45.558	51.179	24
25	10.520	13.120	16.473	24.337	34.382	37.652	40.646	44.314	46.928	52.620	25
26	11.160	13.844	17.292	25.336	35.563	38.885	41.923	45.642	48.290	54.052	26
27	11.808	14.573	18.114	26.336	36.741	40.113	43.194	46.963	49.645	55.476	27
28	12.461	15.308	18.939	27.336	37.916	41.337	44.461	48.278	50.993	56.892	28
29	13.121	16.047	19.768	28.336	39.088	42.557	45.722	49.588	52.336	58.301	29
30	13.787	16.791	20.599	29.336	40.256	43.773	46.979	50.892	53.672	59.703	30
31	14.458	17.539	21.434	30.336	41.422	44.985	48.232	52.191	55.003	61.098	31
32	15.134	18.291	22.271	31.336	42.585	46.194	49.480	53.486	56.329	62.487	32
33	15.815	19.047	23.110	32.336	43.745	47.400	50.725	54.776	57.649	63.870	33
34	16.501	19.806	23.952	33.336	44.903	48.602	51.966	56.061	58.964	65.247	34
35	17.192	20.569	24.797	34.336	46.059	49.802	53.203	57.342	60.275	66.619	35
36	17.887	21.336	25.643	35.336	47.212	50.998	54.437	58.619	61.582	67.985	36
37	18.586	22.106	26.492	36.335	48.363	52.192	55.668	59.892	62.884	69.346	37
38	19.289	22.878	27.343	37.335	49.513	53.384	56.896	61.162	64.182	70.703	38
39	19.996	23.654	28.196	38.335	50.660	54.572	58.120	62.428	65.476	72.055	39
40	20.707	24.433	29.051	39.335	51.805	55.758	59.342	63.691	66.766	73.402	40
41	21.421	25.215	29.907	40.335	52.949	56.942	60.561	64.950	68.053	74.745	41
42	22.138	25.999	30.765	41.335	54.090	58.124	61.777	66.206	69.336	76.084	42
43	22.859	26.785	31.625	42.335	55.230	59.304	62.990	67.459	70.616	77.419	43
44	23.584	27.575	32.487	43.335	56.369	60.481	64.202	68.710	71.893	78.750	44
45	24.311	28.366	33.350	44.335	57.505	61.656	65.410	69.957	73.166	80.077	45
46	25.042	29.160	34.215	45.335	58.641	62.830	66.617	71.201	74.437	81.400	46
47	25.775	29.956	35.081	46.335	59.774	64.001	67.821	72.443	75.704	82.720	47
48	26.511	30.755	35.949	47.335	60.907	65.171	69.023	73.683	76.969	84.037	48
49	27.249	31.555	36.818	48.335	62.038	66.339	70.222	74.919	78.231	85.351	49
50	27.991	32.357	37.689	49.335	63.167	67.505	71.420	76.154	79.490	86.661	50

Figura I.8.1 - Tabela qui-quadrado

Valores Críticos da Distribuição da Estatística D_n (Kolmogorov-Smirnov)

Os valores tabelados correspondem aos pontos $D_{n,\alpha}$ tais que: $P(D_n \geq D_{n,\alpha}) = \alpha$.

α						α					
n	0.20	0.10	0.05	0.02	0.01	n	0.20	0.10	0.05	0.02	0.01
1	0.900	0.95	0.975	0.990	0.995	21	0.226	0.259	0.287	0.321	0.344
2	0.684	0.776	0.842	0.900	0.929	22	0.221	0.253	0.281	0.314	0.337
3	0.565	0.636	0.708	0.785	0.829	23	0.216	0.247	0.275	0.307	0.330
4	0.493	0.565	0.624	0.689	0.734	24	0.212	0.242	0.269	0.301	0.323
5	0.447	0.509	0.563	0.627	0.669	25	0.208	0.238	0.264	0.295	0.317
6	0.410	0.468	0.519	0.577	0.617	26	0.204	0.233	0.259	0.290	0.311
7	0.381	0.436	0.483	0.538	0.576	27	0.200	0.229	0.254	0.284	0.305
8	0.358	0.410	0.454	0.407	0.542	28	0.197	0.225	0.250	0.279	0.300
9	0.339	0.387	0.430	0.480	0.513	29	0.193	0.221	0.246	0.275	0.295
10	0.323	0.369	0.409	0.457	0.489	30	0.190	0.218	0.242	0.270	0.290
11	0.308	0.352	0.391	0.437	0.468	31	0.187	0.214	0.238	0.266	0.285
12	0.296	0.338	0.375	0.419	0.449	32	0.184	0.211	0.234	0.262	0.281
13	0.285	0.325	0.361	0.404	0.432	33	0.182	0.208	0.231	0.258	0.277
14	0.275	0.314	0.349	0.390	0.418	34	0.179	0.205	0.227	0.254	0.273
15	0.266	0.304	0.338	0.377	0.404	35	0.177	0.202	0.224	0.251	0.269
16	0.258	0.295	0.327	0.366	0.392	36	0.174	0.199	0.221	0.247	0.265
17	0.250	0.286	0.318	0.355	0.381	37	0.172	0.196	0.218	0.244	0.262
18	0.244	0.279	0.309	0.346	0.371	38	0.170	0.194	0.215	0.241	0.258
19	0.237	0.271	0.301	0.337	0.361	39	0.168	0.191	0.213	0.238	0.255
20	0.232	0.265	0.294	0.329	0.352	40	0.165	0.189	0.210	0.235	0.252

Para $n > 40$ os valores críticos de D_n podem ser aproximados pelas seguintes expressões:

α				
0.20	0.10	0.05	0.02	0.01
$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Figura I.8.2 - Tabela Kolmogorov-Smirnov

$p =$	0.80	0.85	0.90	0.95	0.99
n= 4	.300	.319	.352	.381	.417
5	.285	.299	.315	.337	.405
6	.265	.277	.294	.319	.364
7	.247	.258	.276	.300	.348
8	.233	.244	.261	.285	.331
9	.223	.233	.249	.271	.311
10	.215	.224	.239	.258	.294
11	.206	.217	.230	.249	.284
12	.199	.212	.223	.242	.275
13	.190	.202	.214	.234	.268
14	.183	.194	.207	.227	.261
15	.177	.187	.201	.220	.257
16	.173	.182	.195	.213	.250
17	.169	.177	.189	.206	.245
18	.166	.173	.184	.200	.239
19	.163	.169	.179	.195	.235
20	.160	.166	.174	.190	.231
25	.142	.147	.158	.173	.200
30	.131	.136	.144	.161	.187
> 30	.736	.768	.805	.886	1.031
	$\frac{\sqrt{n}}$	$\frac{\sqrt{n}}$	$\frac{\sqrt{n}}$	$\frac{\sqrt{n}}$	$\frac{\sqrt{n}}$

Figura I.8.3 - Tabela Lilliefors

Factores para construção das cartas de controlo de variáveis														
Dimensão Amostra	Carta da Média				Carta do Desvios Padrão						Carta da Amplitude			
	Factores Limites Controlo				Factores para Limites de Controlo						Factores para Limites de Controlo			
	A	A ₂	A ₃	d ₂	c ₄	B ₃	B ₄	B ₅	B ₆	d ₃	D ₁	D ₂	D ₃	D ₄
2	2,121	1,881	2,659	1,128	0,7979	0	3,267	0	2,606	0,853	0	3,686	0	3,267
3	1,732	1,023	1,954	1,693	0,8862	0	2,568	0	2,276	0,888	0	4,358	0	2,574
4	1,500	0,729	1,628	2,059	0,9213	0	2,266	0	2,088	0,880	0	4,698	0	2,282
5	1,342	0,577	1,427	2,326	0,9400	0	2,089	0	1,964	0,864	0	4,918	0	2,114
6	1,225	0,483	1,287	2,534	0,9515	0,030	1,970	0,029	1,874	0,848	0	5,078	0	2,004
7	1,134	0,419	1,182	2,704	0,9594	0,118	1,882	0,113	1,806	0,833	0,204	5,204	0,076	1,924
8	1,061	0,373	1,099	2,847	0,9650	0,185	1,815	0,179	1,751	0,820	0,388	5,306	0,136	1,864
9	1,000	0,337	1,032	2,970	0,9693	0,239	1,761	0,232	1,707	0,808	0,547	5,393	0,184	1,816
10	0,949	0,308	0,975	3,078	0,9727	0,284	1,716	0,276	1,669	0,797	0,687	5,469	0,223	1,777
11	0,905	0,285	0,927	3,173	0,9754	0,321	1,679	0,313	1,637	0,787	0,811	5,535	0,256	1,744
12	0,866	0,266	0,886	3,258	0,9776	0,354	1,646	0,346	1,610	0,778	0,922	5,594	0,283	1,717
13	0,832	0,249	0,850	3,336	0,9794	0,382	1,618	0,374	1,585	0,770	1,025	5,647	0,307	1,693
14	0,802	0,235	0,817	3,407	0,9810	0,406	1,594	0,399	1,563	0,763	1,118	5,696	0,328	1,672
15	0,775	0,223	0,789	3,472	0,9823	0,428	1,572	0,421	1,544	0,756	1,203	5,741	0,347	1,653
16	0,750	0,212	0,763	3,532	0,9835	0,448	1,552	0,440	1,526	0,750	1,282	5,782	0,363	1,637
17	0,728	0,203	0,739	3,588	0,9845	0,466	1,534	0,458	1,511	0,744	1,356	5,820	0,378	1,622
18	0,707	0,194	0,718	3,640	0,9854	0,482	1,518	0,475	1,496	0,739	1,424	5,856	0,391	1,608
19	0,688	0,187	0,698	3,689	0,9862	0,497	1,503	0,490	1,483	0,734	1,487	5,891	0,403	1,597
20	0,671	0,180	0,680	3,735	0,9869	0,510	1,490	0,504	1,470	0,729	1,549	5,921	0,415	1,585
21	0,655	0,173	0,663	3,778	0,9876	0,523	1,477	0,516	1,459	0,724	1,605	5,951	0,425	1,575
22	0,640	0,167	0,647	3,819	0,9882	0,534	1,466	0,528	1,448	0,720	1,659	5,979	0,434	1,566
23	0,626	0,162	0,633	3,858	0,9887	0,545	1,455	0,539	1,438	0,716	1,710	6,006	0,443	1,557
24	0,612	0,157	0,619	3,895	0,9892	0,555	1,445	0,549	1,429	0,712	1,759	6,031	0,451	1,548
25	0,600	0,153	0,606	3,931	0,9896	0,565	1,435	0,559	1,420	0,708	1,806	6,056	0,459	1,541

Para $n > 25$

$A = \frac{3}{\sqrt{n}}$
 $A_3 = \frac{3}{c_4 \sqrt{n}}$
 $c_4 \approx \frac{4(n-1)}{4n-3}$
 $B_3 = 1 - \frac{3\sqrt{1-c_4^2}}{c_4}$
 $B_4 = 1 + \frac{3\sqrt{1-c_4^2}}{c_4}$
 $B_5 = c_4 - 3\sqrt{1-c_4^2}$
 $B_6 = c_4 + 3\sqrt{1-c_4^2}$

Figura I.4 - Tabela das constantes das cartas de controlo

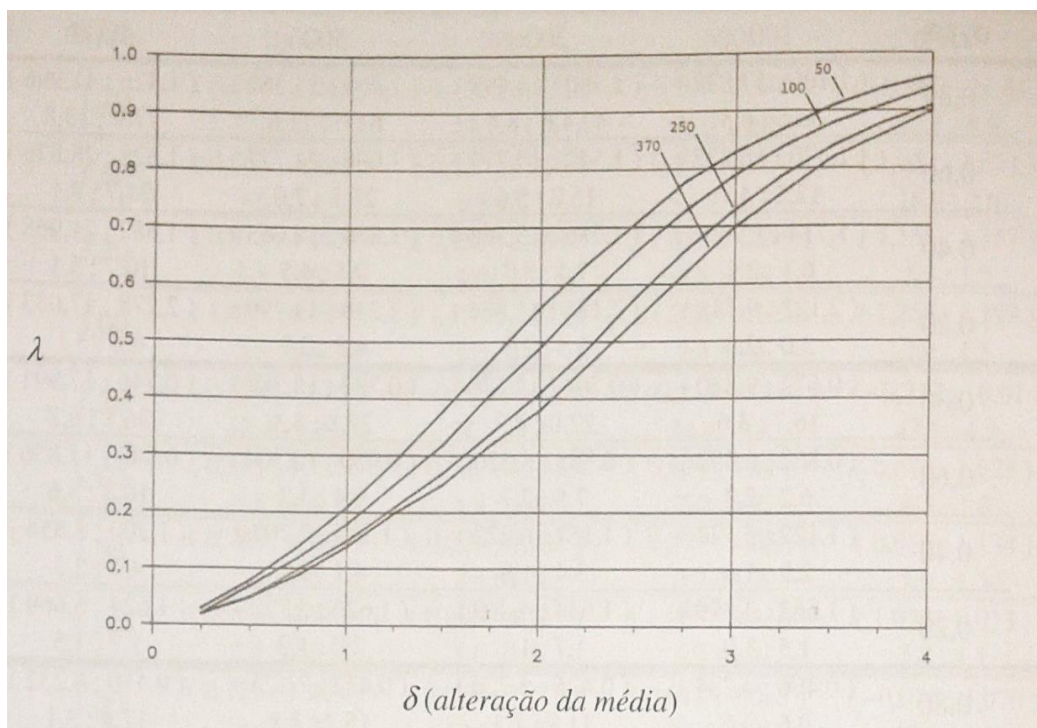


Figura I.5 - Ábaco EWMA para a determinação do parâmetro λ em função do ARL em controle

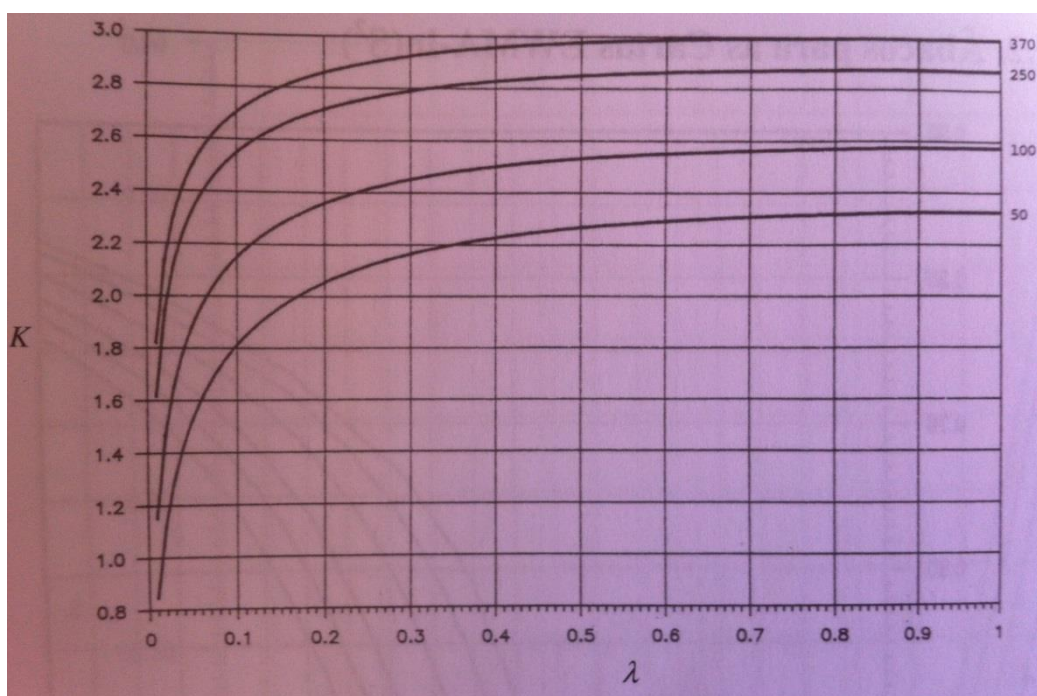


Figura I.6 - Ábaco EWMA para a determinação do parâmetro K em função do ARL em controle

Anexo II – Tratamento de input e outputs

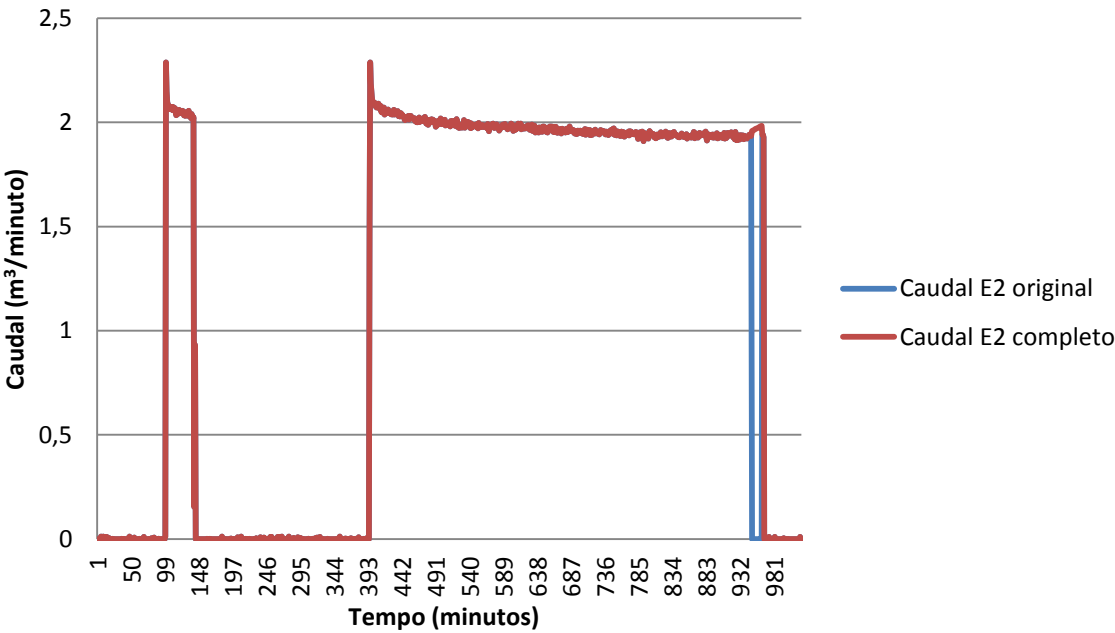


Figura II.1 - Tratamento do input de caudal da entrada E2

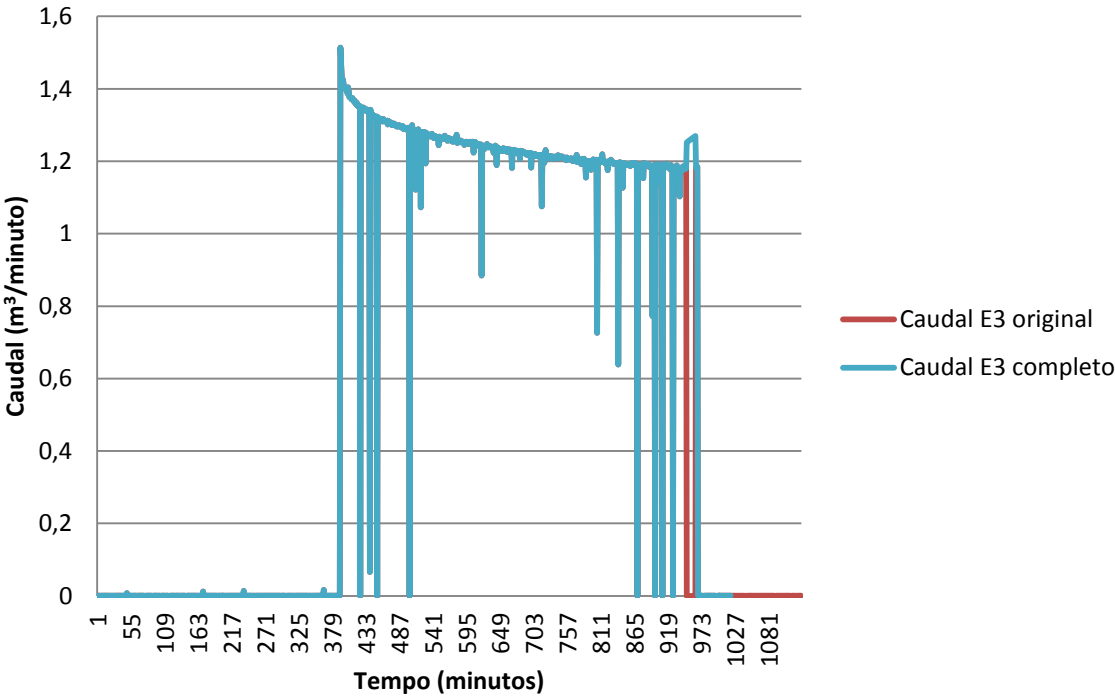


Figura II.2 - Tratamento do input de caudal da entrada E3

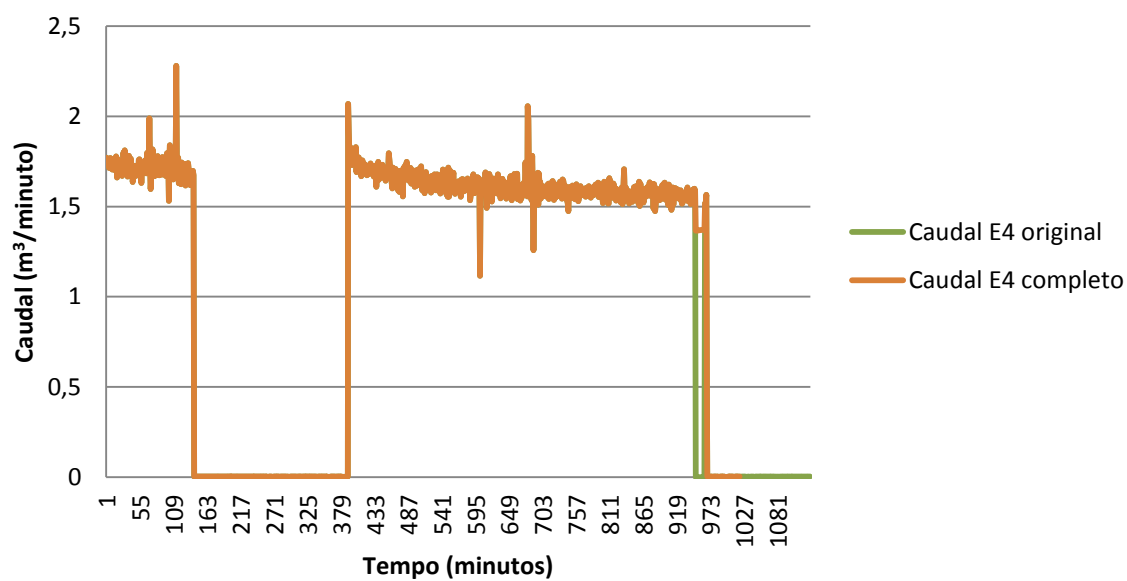


Figura II.3 - Tratamento do input de caudal da entrada E4

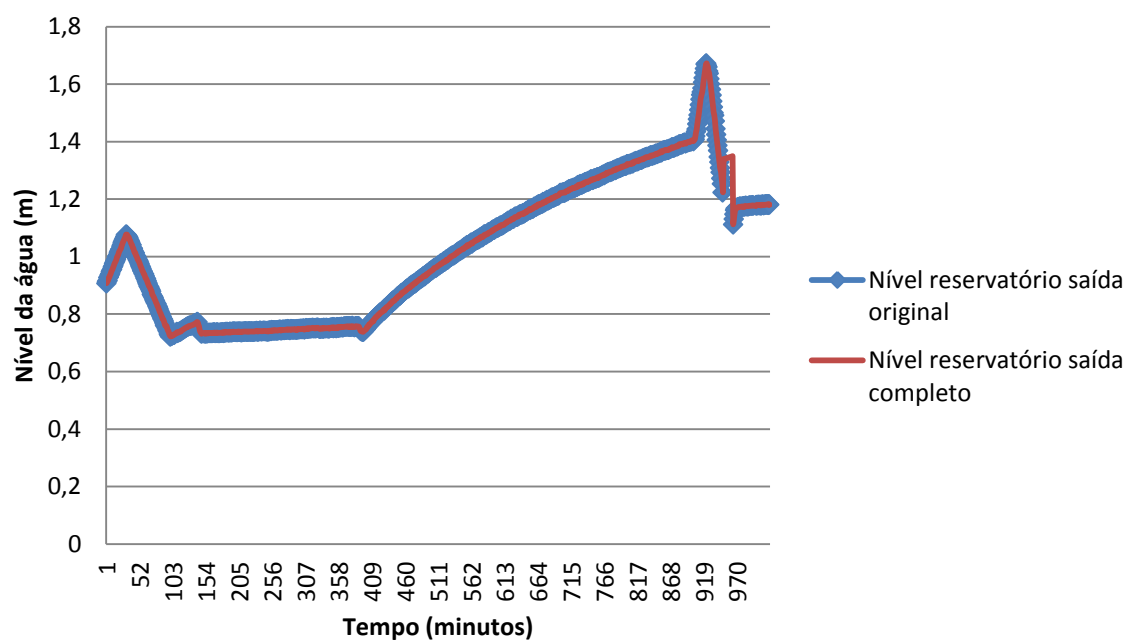


Figura II.4 - Tratamento do input do nível do reservatório de saída

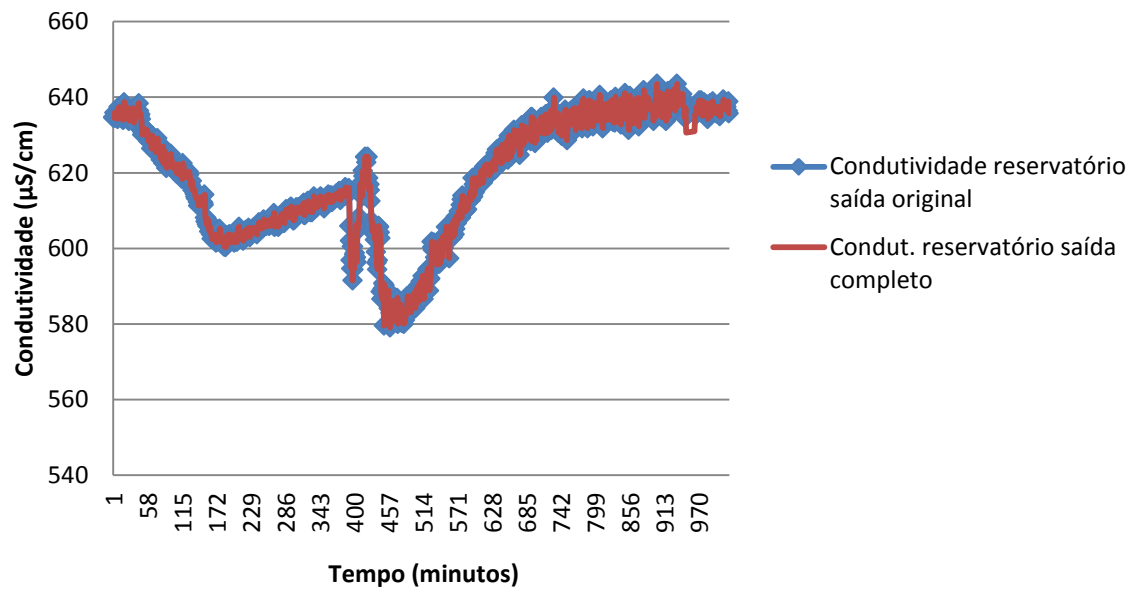


Figura II.5 - Tratamento do input do valor da condutividade do reservatório de saída

Tabela II.1 - Análise de sensibilidade da entrada E4

Entrada E4					
Valor inicial da condutividade	DMA inicial	Alteração 1	DMA Alteração 1	Alteração 2	DMA Alteração 2
508.86	3.23	1400	3.48	1000	3.79

Anexo III – Cartas de controlo e modelo ARIMA

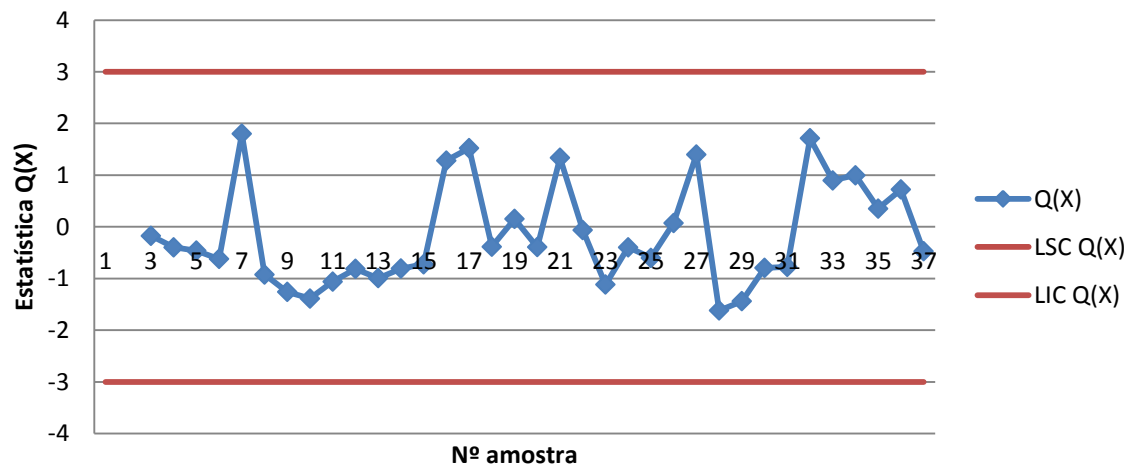


Figura III.1 – Carta Q(X) para o controlo da entrada E3

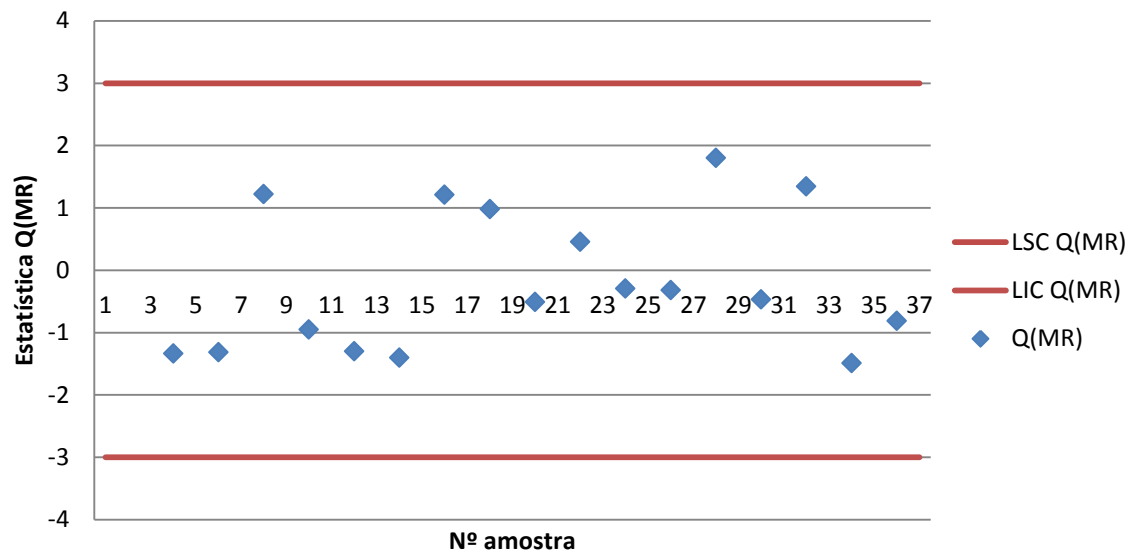


Figura III.2 - Carta Q(MR) para o controlo da entrada E3

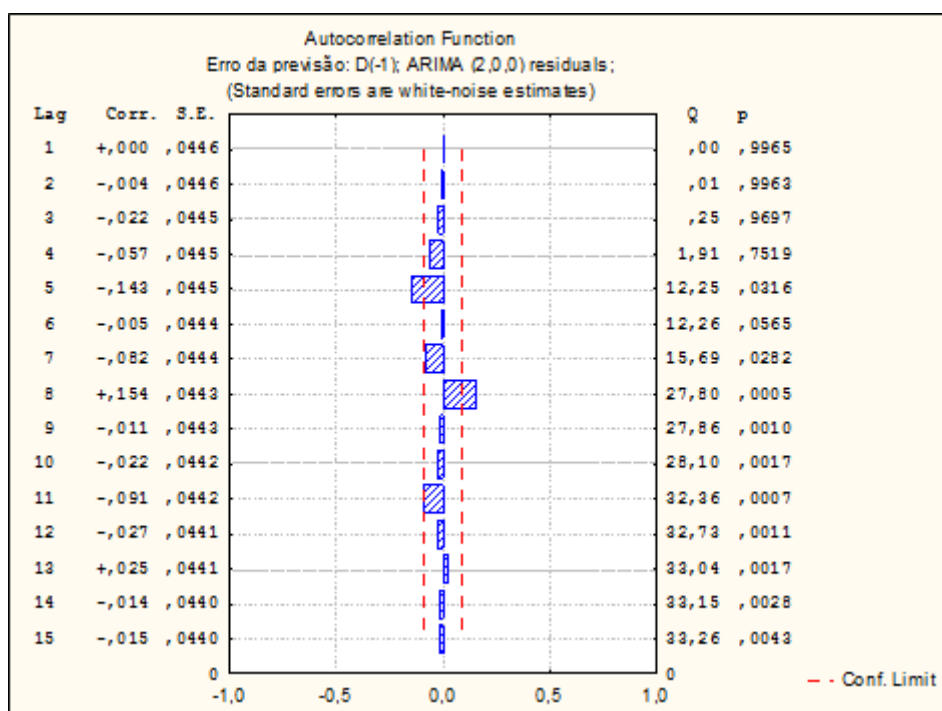


Figura III.3 - Função de autocorrelação do termo erro do modelo ARIMA (2; 1; 0)

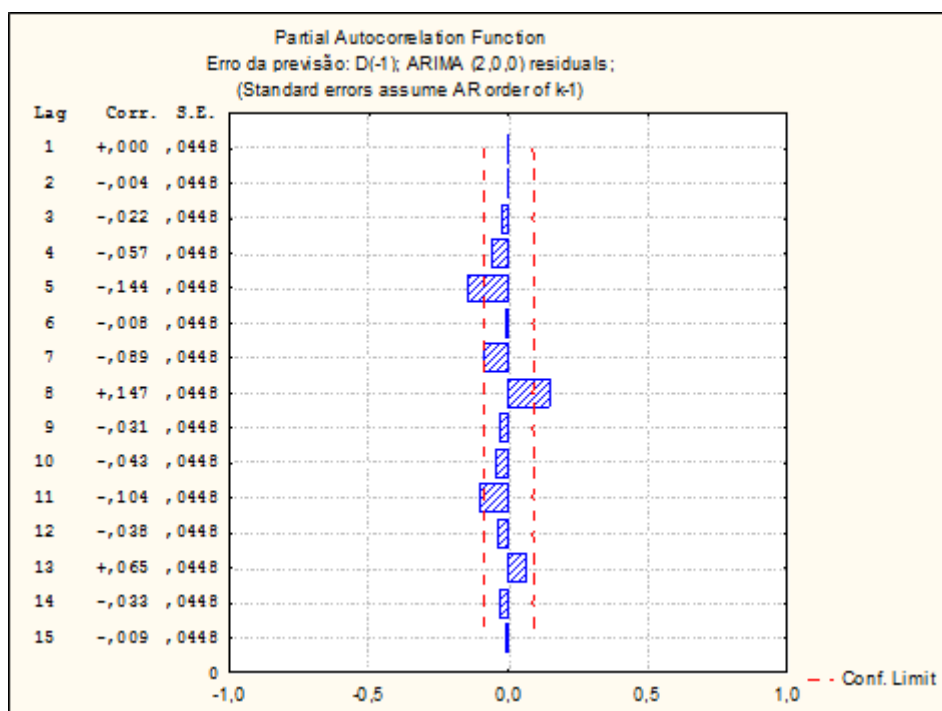


Figura III.4 - Função de autocorrelação parcial do termo erro do modelo ARIMA (2; 1; 0)

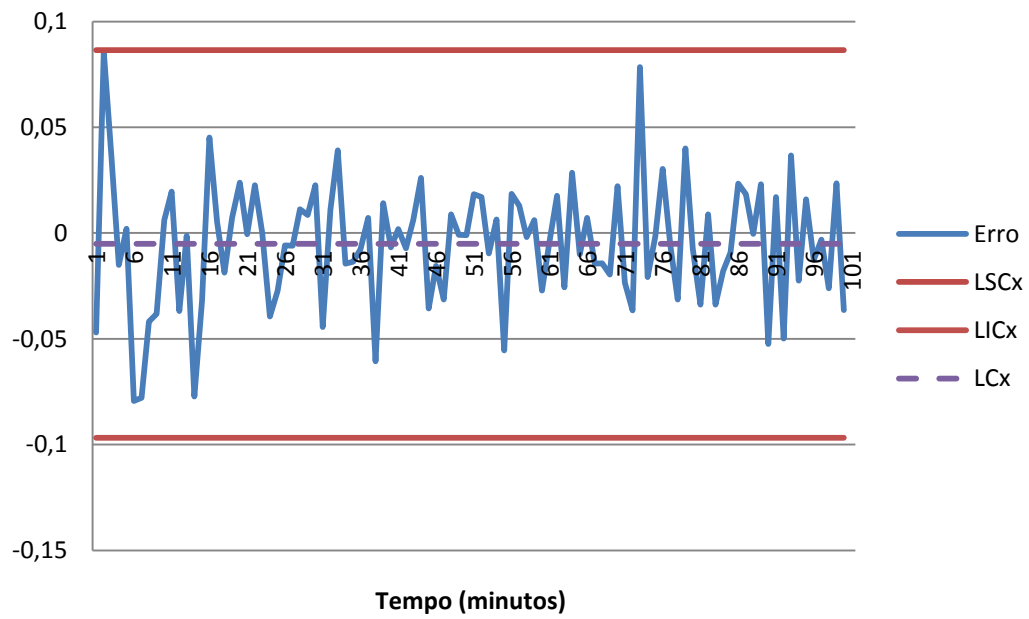


Figura III.5 - Carta X do passo intermédio da aplicação da Fase I ao controlo dos erros do modelo ARIMA

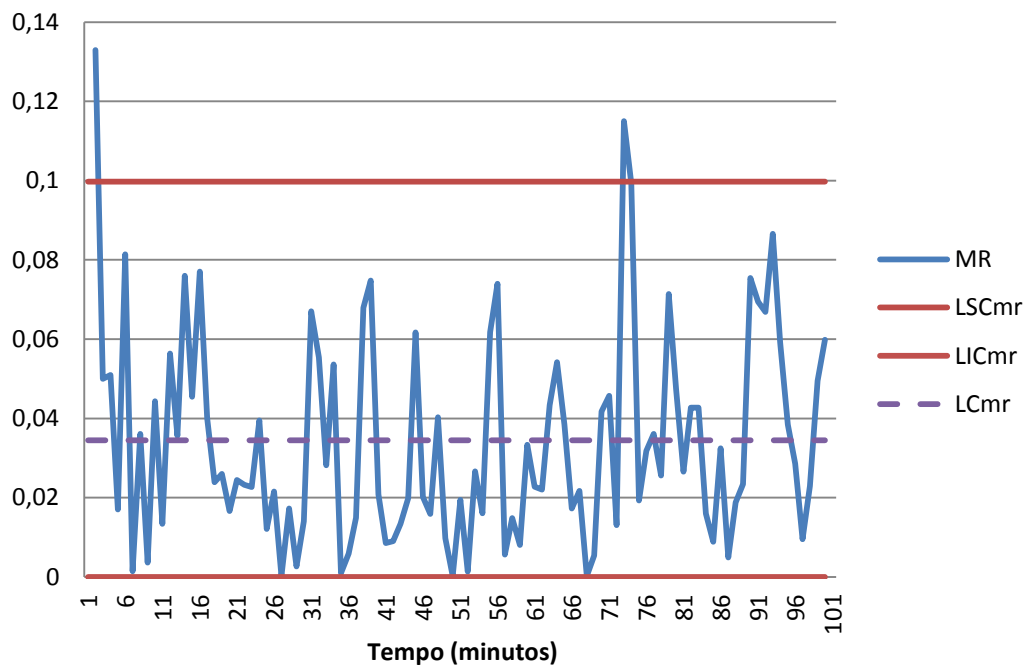


Figura III.6 - Carta MR do passo intermédio da aplicação da Fase I ao controlo dos erros do modelo ARIMA